LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 38
ottobre 1971
anno IV
volume VII

I resti delle supernove

La radiazione emessa dai detriti lasciati da catastrofiche esplosioni stellari può fornire degli indizi su molti enigmi astronomici

di Paul Gorenstein e Wallace Tucker

« Le stelle, come pustole di morbillo, alla fine svaniscono. » Queste parole del poeta Samuel Hoffenstein forniscono una descrizione efficace anche se irriverente del modo in cui le stelle giungono alla fine della loro evoluzione. Alcune stelle, però, quando il loro fuoco nucleare diventa freddo, si rifiutano di fare una fine cosí tranquilla. Circa una stella su 100 esplode in modo catastrofico. Per circa due settimane una stella in esplosione – una supernova – irraggia piú energia di un miliardo di soli e la materia viene proiettata a una velocità che è circa quella della luce. Raggi cosmici (nuclei di idrogeno e di altri elementi leggeri) vengono probabilmente prodotti in abbondanza e si pensa che durante i primi pochi secondi dell'esplosione si sintetizzano elementi piú pesanti del ferro.

Lo strato di detriti che si sta espandendo crea una nebulosa che, per centinaia o persino migliaia di anni, irraggia fortemente sia nella regione radio dello spettro che in quella dei raggi X. Inoltre il cataclisma può lasciarsi dietro un piccolo nucleo estremamente denso. La recente scoperta dei pulsar ha fatto nascere l'ipotesi che questo nucleo in rotazione ad alta velocità possa essere una sorgente pulsante di radiazione. Pertanto la « morte » di una stella è talvolta seguita da una « trasfigurazione » in cui la stella inizia una nuova vita con condizioni fisiche completamente differenti da quelle che si incontrano in una stella normale.

Nella nostra galassia, che contiene qualcosa come 100 miliardi di stelle, si ha l'esplosione di una supernova circa una volta ogni 100 anni. Poiché la nebulosa e la stella « trasfigurata » sopravvivono per lungo tempo dopo lo scoppio, nella nostra galassia ci sarà sempre una notevole quantità di resti di supernove, accessibili perciò a una osservazione dettagliata. Finora ne sono state individuate almeno due dozzine.

Quattro di questi resti di supernove sono anche stati identificati con delle « nove » della cui improvvisa apparizione nel cielo si può trovare notizia nei documenti storici. Questi quattro sono gli oggetti osservati nel 1006, nel 1054 (Nebulosa del Granchio), nel 1572 (nova di Tyge) e nel 1604 (nova di Keplero) – tutti prima dell'invenzione del telescopio. Sebbene le prime due esplosioni fossero probabilmente piú luminose delle ultime due, non c'è nessuna prova che esse abbiano fatto la benché minima impressione sugli abitanti dell'Europa, che erano convinti dell'immutabilità dei cieli. La supernova del 1572 fu tuttavia subito avvistata da Tyge Brahe, che la osservò regolarmente fino a quando scomparve dalla vista; Tyge pubblicò le sue osservazioni nel *De nova... stella* del 1573.

La ricerca dei resti di tale evento non diede risultati fino al 1952, quando R. Hanbury Brown e Cyril Hazard dell'Università di Cambridge scoprirono una intensa sorgente radio vicino ma non proprio nella posizione data da Tyge. Con l'aiuto di questa informazione venne trovato anche il residuo ottico da Rudolph Minkowski con il telescopio Hale da 5 metri. Recenti osservazioni con strumenti installati su razzi, hanno mostrato che i resti della nova di Tyge costituiscono anche una intensa sorgente di raggi X con energie tra i 1000 e 10 000 elettronvolt.

La supernova del 1604 venne studiata da Keplero; essa raggiunse una luminosità uguale a quella di Giove e forse anche un po' superiore. Fu visibile di giorno per circa un mese e di notte per circa un anno. Nel 1941 Walter Baade dell'Osservatorio di Monte Wilson, che aveva analizzato i vecchi dati, scoprí una nebulosa ottica vicino al punto in cui nel 1604 era comparsa la supernova. La nebulosa è formata da filamenti luminosi che coprono una area che sottende circa un minuto d'arco. Circa 10 anni dopo Hambury Brown e Hazard scoprirono che anche essa era una sorgente radio.

Le supernove del 1006 e del 1054 si verificarono in un periodo in cui l'astronomia era molto fiorente in Oriente. In quei giorni l'apparizione di un nuovo oggetto luminoso nel cielo fu un evento portentoso sia per i singoli che per le nazioni, e accurati dati astronomici furono conservati dagli astrologi di corte per poterli interpretare. Nel lontano 1920 Edwin P. Hubble di Monte Wilson fece l'ipotesi che la Nebulosa del Granchio fosse il residuo della nova del 1054. Egli osservò che, dalle dimensioni della nebulosa e dalla velocità di espansione, sembrava che essa avesse avuto origine circa 900 anni pri-

ma. Studi dettagliati di K. Lundmark, Jan H. Oort, Nicholas U. Mayall e Duyvendak confermarono successivamente l'ipotesi di Hubble.

La supernova del 1006 fu individuata in modo definitivo principalmente grazie a uno studio di antiche cronache arabe, siriache, latine, giapponesi e cinesi fatto da Bernard Goldstein. Da esse si era appreso che nell'aprile del 1006 era apparsa una supernova nella costellazione del Lupo. Inizialmente la stella aveva la stessa luminosità di Venere e rimase visibile, di notte, per molto piú di un anno. Oggi è difficile individuare i resti di tale evento nella parte dello spettro del visibile. Tuttavia nel 1965 fu scoperta un'emissione radio proveniente da tali resti.

Non tutte le « stelle strane » di cui è riferito nelle cronache cinesi o le nuove stelle citate dagli astronomi europei sono classificate come supernove. In alcuni casi si tratta di comete, in altri di ordinarie nove: esplosioni negli strati esterni di una stella, molto piú frequenti di una supernova ma meno spettacolari. Non è facile fare una distinzione fra nove e supernove anche se le supernove sono intrinsecamente molto piú luminose. Una nova, se è vicina, può apparire luminosa quanto o anche piú di una supernova che sia molto piú distante. Inoltre la luce emessa da un oggetto distante viene spesso assorbita o diffusa da particelle di polvere nel mezzo interstellare. Per questo motivo fino al 1920 non si scoprí che le supernove costituivano una diversa classe di stelle in esplosione, intrinsecamente molto piú luminose delle nove.

La distinzione fu fatta dopo aver riconosciuto che molti oggetti chiamati nebulose e che si pensava fossero vicini erano in effetti galassie distanti. Nel periodo che va dal 1885 al 1920 vennero scoperte in tali « nebulose » circa 10 nove. In alcuni casi la luminosità della nova era uguale o superiore a quella della nebulosa a essa associata. Quando Hubble dimostrò che nella maggior parte dei casi le nebulose erano sistemi di stelle indipendenti confrontabili con la nostra galassia e distanti piú di un milione di anni luce, divenne chiaro che una nova osservata in un tale sistema doveva avere una luminosità 10 000 volte superiore a quella di una nova ordinaria; esplosioni di quest'ultimo tipo avvengono nella nostra galassia con la frequenza di circa 20 volte all'anno. Fritz Zwicky di Monte Wilson e Baade suggerirono di chiamare tali esplosioni piú catastrofiche « supernove ».

Poiché l'esplosione di una supernova è un evento raro in qualsiasi galassia, l'unico modo per poterne studiare le fasi iniziali è di tenere sotto controllo piú galassie aumentando perciò la probabilità di osservare una supernova. Nel 1933 Zwicky iniziò una ricerca sistematica di esplosioni di supernove in altre galassie. Oggi tale lavoro pionieristico è portato avanti da astronomi in tutto il mondo e vengono scoperte da 10 a 20 esplosioni di supernove all'anno. Tenendo conto del numero di galassie osservate si può arrivare a pensare che, per ogni galassia, si abbia la esplosione di una supernova ogni 100 anni.

Ciò che caratterizza i resti di supernova è l'emissione di onde radio la cui distribuzione energetica in funzione della lunghezza d'onda è « non termica », cioè di tipo non associato a un ordinario oggetto caldo. Anche se sembra che la radiazione abbia origine da una struttura a strati o provenga da una regione che mostra dei filamenti visibili otticamente, la sua sorgente si trova quasi certamente nei resti di una supernova. Probabilmente tutti i resti individuati in tal modo hanno meno di 100 000 anni di vita. I membri piú giovani del gruppo, la cui emissione di energia è 1000 volte quella del Sole, sono tra gli oggetti piú luminosi della nostra galassia. Non tutta l'energia è emessa sotto forma di luce visibile o di onde radio. Nel caso di alcuni resti, molta, se non la maggior parte, dell'energia è emessa sotto forma di fotoni energetici: raggi X con energie che vanno da poche centinaia di elettronvolt a qualcosa come un milione di elettronvolt. Poiché l'energia emessa dai resti di supernove copre un campo di frequenza cosí esteso, dai raggi X alle onde radio, esse hanno in astronomia un ruolo che è unico e sono particolarmente importanti per l'interpretazione dei processi astrofisici.

All'epoca in cui Hubble fece l'ipotesi che la Nebulosa del Granchio fosse ciò che restava della supernova del 1054, gli astronomi avevano pochi strumenti per studiare qualsiasi cosa che non fosse lo spettro visibile. L'atmosfera terrestre blocca radiazioni di lunghezze d'onda piú corte di 3100 angstrom circa, che nello spettro si trovano dalla parte dell'ultravioletto vicino. Sebbene alcune lunghezze d'onda radio penetrino liberamente nell'atmosfera, la radioastronomia prima della seconda guerra mondiale era limitata da una tecnologia inadeguata e gli esperti erano pochi. I progressi fatti in tempo di guerra allargarono notevolmente gli orizzonti dell'astronomia. Essi fecero conoscere tecniche sensibili per la rivelazione di onde radio a una nuova generazione di astronomi che le applicarono con entusiasmo all'osservazione del cosmo. Per quanto riguarda l'altro estremo dello spettro, i progressi del dopoguerra in missilistica permisero, per la prima volta, di innalzare degli strumenti fin sopra la coltre opaca dell'atmosfera terrestre.

Ne risultò che la Nebulosa del Granchio era una delle piú intense sorgenti di onde radio e fu la prima sorgente radio a essere identificata con un oggetto visibile. Si era perciò sperato che le osservazioni radio potessero spiegare alcuni degli aspetti enigmatici della nebulosa. Per esempio le precise osservazioni ottiche avevano mostrato che la velocità di espansione rilevata non poteva portare alle dimensioni osservate a meno che non ci fosse stato, dopo l'esplosione del 1054, un qualche altro tipo di accelerazione. Inoltre erano state scoperte vicino al centro della nebulosa delle configurazioni chiare variabili, delle specie di « increspature ». Le increspature apparivano a intervalli regolari, si muovevano rapidamente verso l'esterno a una velocità che arrivava fino a un decimo di quella della luce e dopo circa un mese scomparivano.

La scoperta che la Nebulosa del Granchio era una potente emettitrice di onde radio approfondí all'inizio soltanto il mistero. Gli astronomi non riuscivano a capire come flussi di emissione radio cosí grandi potessero essere prodotti da un oggetto astronomico. Successivamente gli svedesi Hannes Alfvén e A. Herlofson, il tedesco K. O. Kiepenheuer e il sovietico V. L. Ginzburg mostrarono che ci si poteva aspettare una emissione radio non termica se l'oggetto conteneva due ingredienti: elettroni di energia molto elevata e un campo magnetico con cui essi potessero interagire. Gli elettroni, se intrappolati dal campo, dovrebbero muoversi su orbite attorno alle linee di forza magnetica ed emettere uno stretto raggio di radiazione tangente alle loro orbite e in direzione del moto. La radiazione, comunemente chiamata « radiazione di sincrotrone », poiché fu studiata in connessione con gli elettrosincrotroni acceleratori di particelle il cui sviluppo si ebbe nel 1940, è polarizzata, con il suo vettore elettrico che risulta perpendicolare sia al campo magnetico sia alla direzione di irraggiamento (*si veda la figura a pagina 18*). L'intensità di questa radiazione e la frequenza a cui si ha un picco aumentano rapidamente con l'energia.

Nel 1953 il russo I.S. Shklovsky spiegò come il meccanismo di sincrotrone potesse anche rendere conto dello spettro ottico continuo della Nebulosa del Granchio. In seguito alla os-

L'espansione della Nebulosa del Granchio è chiaramente mostrata in questa fotografia composita ottenuta sovrapponendo una immagine positiva e una negativa eseguite a 14 anni di distanza. Le lastre originali sono state entrambe scattate con il telescopio Hale da 5 metri a Monte Palomar; il positivo, in cui i filamenti gassosi della nebulosa appaiono chiari, è stato stampato da una lastra fatta nel 1950 da Walter Baade; il negativo, in cui la stessa struttura appare scura, è stato stampato da una lastra fatta nel 1964 da Guido Münch a cui si deve anche la composizione. Il lieve spostamento relativo della immagine chiara e di quella scura, dovuto all'espansione della nebulosa verificatosi tra le due esposizioni, fa sí che la fotografia composita abbia l'aspetto di un bassorilievo. La Nebulosa del Granchio è ciò che resta dell'esplosione di una supernova osservata nel 1054 dagli astronomi cinesi. Si trova alla distanza di circa 5000 anni luce e la porzione attualmente visibile misura circa sei anni luce. Questa nebulosa è anche una potente sorgente di onde radio a raggi X. Tale prodigiosa produzione di energia è stata attribuita alla trasformazione dell'energia cinetica della stella centrale rotante (*freccia*) in energia elettromagnetica.

L'intera struttura del Nodo del Cigno è stata fotografata in luce rossa con il telescopio Schmidt da 120 centimetri. L'espansione di questa struttura del tipo a strati si sta probabilmente verificando da ben 60 000 anni; attualmente ha un diametro di circa 100 anni luce. Questo residuo di supernova risulta essere anche una intensa sorgente sia di onde radio sia di raggi X.

servazione successiva che l'emissione ottica e radio sono polarizzate, venne comunemente accettata l'ipotesi del sincrotrone (*si veda la figura a pagina 19*). Divenne così chiaro che la Nebulosa del Granchio è un « sincrotrone cosmico » pieno di elettroni con energie di 1000 miliardi di elettronvolt o anche più elevate, e di campi magnetici più di 10 volte più intensi del campo dello spazio interstellare. Il problema di capire quali siano le condizioni della Nebulosa del Granchio o di altri resti di supernova, divenne quello di capire la origine del campo magnetico e quella degli enormi flussi di elettroni ad alta energia.

Un nuovo elemento si aggiunse alla storia nel 1964 quando gli strumenti situati a bordo di razzi sonda, rivelarono emissione di raggi X dalla Nebulosa del Granchio. Sebbene anche i raggi X possano essere prodotti da un processo di sincrotrone, questa interpretazione presentava delle difficoltà poiché elettroni di sincrotrone sufficientemente energetici per produrre raggi X perderebbero rapidamente la loro energia per irraggiamento. Nacque perciò il problema di come ci potesse essere un continuo rifornimento di elettroni di energia sufficiente, circa 900 anni dopo l'esplosione iniziale.

Successivamente nel 1967 furono scoperti i pulsar. Questi oggetti eccezionali emettono brevi impulsi di onde radio a intervalli estremamente regolari che vanno da circa un quarto di secondo a un po' più di un secondo. La loro identità fu chiarita l'anno successivo quando vennero scoperti dei pulsar con periodi molti più brevi nella Nebulosa del Granchio e in un altro residuo di supernova: Vela X. Si stabilì così una profonda connessione tra pulsar e resti di supernova. Inoltre, gli impulsi della Nebulosa del Granchio si potevano osservare sia nella regione visibile dello spettro sia in quella dei raggi X. La sorgente di questi impulsi è un oggetto debole che nelle fotografie astronomiche sembra una stella ordinaria. Recentemente si è scoperto che per raggi X di energia superiore a qualche milione di elettronvolt l'emissione di energia della componente pulsante è confrontabile con l'energia della radiazione non pulsata emessa dalla intera nebulosa.

La più recente esplosione di supernova nella nostra galassia si verificò attorno al 1700 nella costellazione di Cassiopea non lontano dalla supernova di Tyge. Può sembrare sorprendente che questo evento non fosse stato osservato sebbene a quell'epoca vi fosse un interesse piuttosto notevole per la astronomia. La spiegazione sembra stia nel fatto che la luce proveniente da quella regione del cielo è oscurata da una grande quantità di polvere; inoltre può darsi che la supernova stessa sia stata meno luminosa delle supernove del 1572 e del 1604. Il primo indizio che in quella regione ci fossero i resti di una supernova venne nel 1948 da Martin Ryle e F.G. Smith che, con l'aiuto di uno dei primi radiotelescopi dell'Università di Cambridge, scoprirono in Cassiopea una intensa sorgente

In questa figura sono presentati altri aspetti del Nodo del Cigno. Quella in alto è una mappa dell'emissione radio proveniente dalla regione della nebulosa fatta di S. Kenderdine dell'Università di Cambridge; i numeri sulle linee indicano l'intensità radio in multipli di 900 gradi Kelvin. L'area rettangolare nera corrisponde alla zona della fotografia nella pagina a fronte; le parti tratteggiate indicano le principali caratteristiche visibili della nebulosa. La lunga freccia bianca indica la direzione esaminata da un rivelatore di raggi X durante il volo di un razzo in cui venne osservata l'emissione di raggi X dalla regione della nebulosa. Il grafico in basso mostra la registrazione dell'intensità dei raggi X integrata su intervalli lungo la direzione di osservazione.

radio che essi chiamarono Cassiopea A. In seguito col telescopio Hale da 5 metri Baade e Minkowski riuscirono a ottenere una fotografia di una debole nebulosa a frammenti che poteva essere identificata con la sorgente radio. In base a studi sulla nebulosa si poté stabilire l'epoca dell'evento.

Negli ultimi anni, grazie a strumenti installati su razzi, si è scoperta una intensa emissione di raggi X da Cassiopea A. La struttura del residuo, che meglio si rivela in una mappa della sua emissione radio, consiste in uno strato in espansione il cui diametro è di circa quattro minuti d'arco *(si vedano le figure a pag. 20 e a pag. 21)*. La struttura a strati fa pensare che la materia sia proiettata fuori ad alta velocità. Per ora la ricerca di un oggetto centrale o di una sorgente di radiazione pulsante non ha avuto successo.

Un esempio di resti di supernova molto piú antichi è dato dal « Nodo del Cigno » *(si veda la figura a pagina 16)*. In confronto alla Nebulosa del Granchio e a Cassiopea A, è un oggetto enorme che si estende nel cielo per tre gradi d'arco e ha un diametro che è sei volte il diametro apparente della Luna. Come nel caso di Cassiopea A si tratta di uno strato di espansione, ma l'espansione è in atto da molto piú tempo, può darsi anche da 60 000 anni. Il Nodo del Cigno è anche una sorgente intensa di onde radio non termiche e l'anno scorso si scoprí che era anche una intensa sorgente di raggi X. Tuttavia i raggi X provenienti dal Nodo del Cigno hanno una energia media piú bassa di quella dei raggi X emessi dalla Nebulosa del Granchio. La ricerca di una stella centrale o di un pulsar nel Nodo del Cigno non ha avuto successo.

Ancora oggi la causa di un'esplosione di supernova è oggetto di intensa ricerca e anche di notevoli controversie; si è però generalmente d'accordo nel ritenere che l'inizio dell'esplosione sia in definitiva collegato all'instabilità nella struttura della stella che si raggiunge quando si esaurisce il rifornimento di combustibile nucleare al centro della stella. Questo stato di instabilità si ha solo in stelle la cui massa sia molto piú grande di una volta e mezzo circa quella del Sole. Stelle di massa minore, fra cui il Sole, incominciano a contrarsi quando è finito il loro combustibile nucleare. Alla fine la forza di gravità è equilibrata dalla pressione degli elettroni « degenerati »: un fluido di elettroni incompressibile che si forma alla fine, poiché due elettroni non possono occupare lo stesso stato energetico. Quando è raggiunta questa configurazione stabile, la stella è diventata una nana bianca e lentamente muore « non esplodendo ma piagnucolando ». Nel caso di stelle la cui massa sia superiore a circa 1,5 masse solari, la densità e la temperatura nel nucleo centrale superano i valori critici oltre i quali non è possibile la stabilità. La stella subisce allora il collasso sotto la azione della forza di gravità e ne risulta un'esplosione.

Nell'esplosione di una supernova gli strati esterni della stella che subisce il collasso vengono proiettati fuori ad alta velocità; in alcuni casi, se non in tutti, viene lasciato dietro un relitto denso. L'idea che si possa avere la formazione di corpi superdensi come prodotti finali di esplosioni di supernove, non è nuova.

L'idea di ipotetiche stelle superdense composte quasi totalmente di neutroni era già stata concepita circa 40 anni fa indipendentemente dai fisici L.D. Landau e Robert Oppenheimer. Nel 1938 Zwicky fece l'ipotesi che tali stelle si potessero formare in esplosioni di supernova. Il collasso gravitazionale che ha luogo prima dell'esplosione di supernova comprime la materia nel nucleo della stella fino a fargli raggiungere un diametro di circa 10 chilometri. Poiché quel che resta del nucleo è ancora una massa confrontabile con quella del Sole, la sua densità è circa uguale alla densità della materia nel nucleo atomico: dell'ordine di 100 000 tonnellate al millimetro cubo.

In seguito alla scoperta dei pulsar l'esistenza di stelle di neutroni è stata largamente accettata. Inoltre i fisici teorici hanno previsto che se una stella con una massa molto grande esplodesse come una supernova, il nucleo potrebbe continuare a contrarsi gravitazionalmente senza limite fino a formare quello che è stato chiamato un « buco nero ». Il campo gravitazionale di un buco nero sarebbe cosí intenso da non lasciar sfuggire nulla dalla sua stretta, neppure la luce. Se i buchi neri esistono essi possono essere rivelati solo per mezzo di effetti gravitazionali. Per esempio, un buco nero potrebbe perturbare il moto di un suo compagno visibile o interagire con la sua radiazione in modo osservabile.

Il collasso gravitazionale che crea una stella di neutroni porta necessariamente a un aumento dell'intensità del campo gravitazionale della stella e della sua velocità di rotazione. Si può immaginare che le linee di flusso magnetico che passano attraverso un materiale altamente conduttore quale è la materia all'interno di una stella siano « congelate » nel materiale; quando il conduttore si muove il campo magnetico si muove con esso. Pertanto quando il nucleo di una stella subisce il collasso gravitazionale, anche il campo magnetico viene compresso assieme alla materia. Il risultato è che il campo magnetico di una stella di neutroni può essere oltre mille miliardi di volte piú intenso del campo magnetico medio del Sole.

La rapidità della rotazione di una stella di neutroni deriva anch'essa dal modo in cui si è formata la stella. Quasi tutte le stelle ruotano e il periodo di rotazione va da qualche ora fino a circa un mese. Il principio di conservazione della quantità di moto comporta che se una stella in rotazione si contrae essa deve ruotare piú velocemente a meno che il momento della quantità di moto (momento angolare) non sia trasferito ai campi di radiazione e ai detriti in espansione. Poiché il raggio del nucleo si riduce di un fattore 10 000 o anche piú quando una stella subisce il collasso fino a raggiungere la configurazione di una stella di neutroni, ne deriverebbe un periodo di rotazione di un solo millisecondo. Note le dimensioni e il periodo di rotazione di una stella di neutroni e assumendo che si tratti di un corpo rigido, valutare l'energia accumulata nella rotazione è una cosa semplice. Per un periodo di 10 millisecondi l'energia rotazionale di una stella di neutroni è dell'ordine di 10^{50} erg, una quantità che è comparabile all'energia irradiata da una stella durante tutta la sua esistenza.

Noi vediamo perciò che un'esplosione di supernova libera energia gravitazionale sotto molte forme. Innanzi tutto c'è l'energia irraggiata direttamente nelle prime fasi dell'esplosione. La materia che contemporaneamente è proiettata fuori a grande velocità, si porta via energia cinetica traslazionale. Infine la stella di neutroni che sopravvive è fornita di un'enorme quantità di energia cinetica rotazionale. Sebbene le nostre conoscenze non siano complete, sembra che si possa rendere conto delle caratteristiche osservate nei resti di supernova mediante trasformazione, entro un periodo di tempo relativamente lungo, di energia cinetica rotazionale e traslazionale in radiazione elettromagnetica.

Vediamo ora come queste idee possano applicarsi al residuo della « stella strana » del 1054, la Nebulosa del Granchio. La scoperta di un pulsar nella Nebulosa del Granchio con un periodo di soli 30 millisecondi fece accettare in generale l'idea che i pulsar siano stelle di neutroni rotanti. Quando un gruppo di radioastronomi all'Osservatorio di Arecibo a Puerto Rico trovò che la velocità di rotazione del pulsar diminuisce in un anno di una parte su 2400, divenne chiaro che c'era una connessione energetica tra il pulsar e la nebulosa. Questo rallentamento, sebbene apparentemente piccolo, implica che il pulsar del Granchio liberi energia rotazionale a una velocità tale da mantenere costante l'energia emessa da tutta la nebulosa. È stata cosí finalmente scoperta la sorgente di energia, cercata a lungo, che alimenta la Nebulosa del Granchio. Il problema ora è quello di capire come dell'energia rotazionale si possa trasferire a particelle molto energetiche. Si è generalmente d'accordo nel pensare che la trasformazione avvenga tramite processi che hanno luogo in un intenso campo magnetico in rapida rotazione. La rapida rotazione e gli intensi campi magnetici contribuiscono a creare vasti campi elettrici che accelerano le particelle fino a energie elevate. Le particelle producono gli impulsi osservati di radiazione ottica, radio e di raggi X. Le particelle, riversandosi fuori nella nebulosa, creano una sorgente continua ed estesa di emissione radio, ottica e di raggi X. Unitamente alle onde elettromagnetiche estese, le particelle accelerate portano via energia e momento angolare della stella di neutroni facendola cosí rallentare.

Questa descrizione generale vale probabilmente anche per quanto riguarda l'attività osservata negli altri resti di supernove recenti. Non si può avere la stessa certezza della Nebulosa del Granchio, poiché fino a oggi si è potuto osservare un solo pulsar in un altro residuo: Vela X. È probabile che ci siano dei pulsar in altri resti di supernova ma che siano orientati in modo da non

La radiazione di sincrotrone è emessa da elettroni che percorrono delle spirali a velocità relativistica (cioè a una velocità che si avvicina a quella della luce) lungo una linea di forza in un campo magnetico. La radiazione viene emessa in uno stretto raggio lungo la direzione di moto degli elettroni; la lunghezza d'onda predominante dipende dalla velocità degli elettroni e dall'intensità del campo magnetico. La radiazione prodotta da questo complesso meccanismo risulta fortemente polarizzata.

La polarizzazione della luce emessa dalla Nebulosa del Granchio è chiaramente mostrata in questa serie di fotografie fatte con un filtro polarizzatore in quattro diversi orientamenti. La freccia nell'angolo in basso a destra di ogni fotografia indica la direzione del vettore elettrico della luce che passa attraverso il filtro. Queste fotografie possono essere considerate la prova che la luce dalla Nebulosa del Granchio è radiazione di sincrotrone emessa da elettroni ad alta energia.

favorire la nostra rivelazione. Secondo una delle possibili interpretazioni dei dati sperimentali, le particelle si riverserebbero fuori lungo uno (o piú) stretto cono da una (o piú) macchia sulla stella di neutroni rotante, cosicché un pulsar assomiglierebbe a un faro. Se questo punto di vista è corretto dovremmo aspettarci di osservare direttamente un pulsar solo se capita che il raggio ruotando spazzi anche il sistema solare. In caso contrario si osserverebbero solo gli effetti indiretti dovuti alle particelle ad alta energia interagenti nella nebulosa. Se l'ampiezza del fascio emesso dal pulsar della Nebulosa del Granchio è normale (meno di 12 gradi), si può facilmente immaginare che soltanto un pulsar su 30 potrà essere rivelato. L'aver osservato dei pulsar soltanto in due resti di supernova che si conoscono è in accordo con questa previsione. Un altro argomento indiretto a favore dell'esistenza di pulsar in tutti i resti di supernova è la necessità di una continua iniezione di particelle ad alta energia nella nebulosa, almeno per qualche anno dopo l'esplosione, per poter giustificare un'emissione radio dell'intensità osservata.

Col passar del tempo la velocità di rotazione della stella di neutroni diminuisce a un punto tale che dalla sorgente non vengono piú praticamente emessi raggi X. L'emissione radio può tuttavia continuare piú a lungo perché può essere mantenuta anche con energia elettronica minore e campi magnetici piú bassi. Quando la sorgente centrale non fornisce piú elettroni ad alta energia, la nebulosità perde anche la sua capacità di produrre raggi X tramite il processo di sincrotrone. L'emissione radio dura molto piú a lungo poiché è prodotta da elettroni di energia piú bassa che possono irraggiare per un periodo molto piú lungo degli elettroni che producono raggi X. È probabile che tutti i circa 50 pulsar noti siano i nuclei rotanti lasciati da esplosioni di supernova, anche se siamo riusciti a rivelare i resti di nebulosa solo per due di essi. Sorge cosí il problema se ogni esplosione di supernova lasci o no dietro di sé un pulsar. Sulla risposta non tutti sono d'accordo. Il numero totale di pulsar che è stato trovato è in accordo, almeno in ordine di grandezza, con una valutazione basata sulla frequenza con cui si verificano esplosioni di supernova, sulla vita media di un pulsar e sull'ipotesi che sia stata rivelata solo quella frazione di pulsar che irraggia nella nostra direzione. Considerando le indeterminazioni in queste quantità non si può escludere la possibilità che soltanto alcune delle esplosioni producano un pulsar; le altre possono lasciare un buco nero o

Cassiopea A: fu identificata per la prima volta nel 1948 come un'intensa radiosorgente da Martin Ryle e F.G. Smith dell'Università di Cambridge. Successivamente Baade e Rudolph Minkowski, che per ottenere delle fotografie come quella di questo negativo fatto in luce rossa da Minkowski usavano il telescopio Hale da 5 metri, mostrarono che corrispondeva a una debole nebulosa a frammenti. Si calcola che l'esplosione di supernova che generò Cassiopea A si sia verificata attorno al 1700.

addirittura nessun oggetto compatto.

Lasciateci ora tentare di prevedere quello che accadrà della materia proiettata dall'esplosione a velocità supersoniche nel mezzo interstellare. Il quadro d'insieme è il seguente: i detriti che si stanno espandendo dal momento dell'esplosione agiscono come un pistone che penetra nel mezzo interstellare e produce un'onda d'urto che viaggia un po' piú veloce della materia espulsa; l'energia della materia espulsa si trasforma quindi gradualmente nell'energia dell'onda d'urto nel mezzo interstellare. Durante le prime migliaia o anche decine di migliaia di anni la temperatura della materia che sta dietro il fronte d'urto è estremamente elevata: da uno a 100 milioni di gradi Kelvin. Un gas a questa temperatura emette la maggior parte della sua energia sotto forma di raggi X e l'intensità dell'emissione dipende dall'entità della massa presente nello strato caldo. Se i resti di supernova si trovano, per esempio, a una distanza di 2000 anni luce, si avrà un flusso rivelabile di raggi X quando si sia ammassata una quantità di materia interstellare pari a circa 25 masse solari, e questo richiede circa 10 000 anni. L'onda d'urto, espandendosi, rallenta e la temperatura della materia dietro il fronte d'urto diminuisce. Di conseguenza verrà emessa radiazione di lunghezza d'onda sempre maggiore. Tuttavia la potenza totale della radiazione aumenta poiché aumenta costantemente la massa di materia interstellare che si ammassa dietro l'onda d'urto. Il processo continua finché la maggior parte dell'energia cinetica trasportata dalla materia originariamente espulsa e poi trasformatasi in energia dell'onda d'urto non sia stata irraggiata, dopo di che l'emissione diminuisce rapidamente. Questo può avvenire in qualsiasi momento, da 30 000 a poche centinaia di migliaia di anni, a seconda dell'energia dell'esplosione e della densità del mezzo interstellare. Durante la fase finale, il ritmo a cui irraggerà l'onda d'urto sarà pari a quello di circa 100 000 soli, soprattutto nella parte dello spettro dell'ultravioletto piú lontano. La temperatura del gas interstellare contenuto in una porzione di spazio piuttosto grande sarà portata dalla radiazione al punto di ionizzazione, punto in cui gli elettroni esterni vengono strappati dagli atomi e dalle molecole.

Un probabile esempio di un'onda di esplosione prodotta da una supernova è il Nodo del Cigno. Il Nodo è una sorgente estesa che emette raggi X, radiazione ottica e radio. Supponendo che si possa applicare a questo esempio il modello dell'onda d'urto, l'emissione di raggi X presumibilmente proviene dal

Mappa radio di Cassiopea A fatta nel 1965 a Cambridge da Ryle e dai suoi collaboratori. L'intervallo tra due linee è di 3000 gradi Kelvin; le zone di grigio progressivamente sempre piú scuro indicano picchi di intensità piú elevata. Il campo visivo è lo stesso della fotografia della pagina a fronte; la posizione delle stelle luminose è indicata nel campo da crocette. La mappa mostra che Cassiopea A ha una struttura circolare che è probabilmente la proiezione di un guscio sferico.

plasma caldo che si trova subito dietro il fronte d'urto e l'emissione ottica dai filamenti piú freddi e piú densi. L'emissione radio è dovuta alla radiazione di sincrotrone che si ha quando elettroni molto veloci incontrano le zone di intensi campi magnetici prodotti dall'onda d'urto. Questi elettroni sono stati probabilmente emessi in un periodo precedente allorché l'energia rotazionale della stella di neutroni o pulsar si era trasferita a particelle che si muovono a una velocità molto prossima a quella della luce.

Possiamo ora riassumere dicendo che, se si forma un pulsar, l'emissione nei primi 1000 anni circa dopo l'esplosione sarà prevalentemente dovuta all'interazione degli elettroni ad alta energia del pulsar con il campo magnetico della nebulosa, con conseguente emissione di raggi X, luce visibile e onde radio. Quando il residuo ha raggiunto un'età di circa 10 000 anni, dietro l'onda d'urto generata dall'esplosione che si sta espandendo si sarà ammassata una grande quantità di materia interstellare che avrà raggiunto una temperatura molto elevata. Questa fase sarà caratterizzata da emissione termica di raggi X e di radiazione ultravioletta. Pertanto, per alcune decine o anche centinaia di migliaia di anni, l'emissione radio e il calore dietro l'onda d'urto indicheranno la posizione in cui nello spazio si è verificata un'esplosione di supernova. Dopo di che, per un osservatore in un piano perpendicolare, gli impulsi radio sempre piú deboli provenienti dalla stella di neutroni potranno continuare per milioni di anni, a ricordo di uno dei piú catastrofici eventi che si abbiano in tutta la natura.

È chiaro che le supernove e i loro resti sono, su scala stellare, molto importanti. Ci si potrebbe domandare se un processo simile a quello che abbiamo descritto possa aver luogo su scala galattica e spiegare forse la tremenda emissione di energia osservata nei quasar, oggetti simili a stelle che sembrano irraggiare un'energia pari a quella irraggiata da 50 o 100 galassie. Resti di supernova e quasar sono, sotto certi aspetti, molto simili. Entrambi, per esempio, sono sorgenti non termiche di onde radio in cui sono continuamente prodotti grandi flussi di particelle ad alta energia, e in entrambe la sorgente fondamentale di energia è concentrata in un piccolo volume. Un'ipotesi è che l'attività di un quasar sia data nientemeno che da una serie di esplosioni di supernove che genera una sorgente piú o meno costante di radiazione.

Un'altra categoria di modelli è basata sull'idea che un oggetto fortemente magnetizzato in rapida rotazione possa essere una efficiente macchina per la produzione di elettroni ad alta energia. Se una stella di neutroni può fornire energia alla Nebulosa del Granchio, e se è possibile che ciò si verifichi anche per gli altri resti, allora un oggetto di massa molto maggiore, rotante nel nucleo di una galassia, potrebbe probabilmente spiegare l'esistenza dei quasar. Secondo questi modelli l'oggetto in rotazione dovrebbe avere una massa di circa un miliardo di masse solari, un diametro dell'ordine di quello del sistema solare, un campo magnetico di intensità tra 1000 e un milione di gauss e un periodo di rotazione che può andare da un anno fino a un giorno, secondo il raggio dell'oggetto ruotante. La sorgente di energia cinetica di rotazione, la rotazione stessa, deriverebbe dal collasso gravitazionale come nel caso di una stella di neutroni. L'energia rotazionale si trasferirebbe a elettroni molto energetici oppure si trasformerebbe in radiazione elettromagnetica grazie agli intensi campi elettrici su grande scala prodotti dal campo magnetico ruotante, analogamente a quanto avviene per il pulsar della Nebulosa del Granchio. Potrebbe quindi accadere che riuscendo a interpretare i resti di supernova si possa risolvere l'enigma dei quasar. Se questo sia vero oppure no resta da vedere, ma in ogni caso i resti di supernova continueranno a servire da straordinari laboratori in cui si possono studiare i processi fondamentali della natura in condizioni fisiche estreme.

Questa successione di disegni illustra l'evoluzione di un residuo di supernova, e inizia con la stella nel suo strato di pre-supernova (*1*). Quando la stella esplode (*2*), lo strato esterno viene proiettato nello spazio, il nucleo subisce un collasso fino a formare una stella di neutroni con un raggio di 10 chilometri e viene emesso un flusso enorme di radiazioni. Nello stadio di supernova in attività (*3*) la stella di neutroni altamente magnetizzata e in rapida rotazione che si trova al centro, produce elettroni relativistici che, muovendosi lungo spirali, entrano nella nebulosa emettendo raggi X, luce visibile e onde radio con meccanismo di sincrotrone. Potrà essere osservata anche della radiazione emessa a impulsi con un periodo uguale a quella della stella di neutroni rotante. La Nebulosa del Granchio è un esempio di residuo di supernova in questo stadio.

Nello stadio successivo dell'evoluzione (*4*) l'energia dello strato in espansione va spesa nella produzione di un'onda d'urto nel mezzo interstellare. Il plasma caldo dietro questa onda d'urto è una sorgente di raggi X e di radiazione nell'ultravioletto piú lontano. Gli elettroni relativistici lasciati dallo stadio attivo, quando vengono a contatto con l'intenso campo magnetico nella regione che si trova dietro l'onda d'urto, producono radiazione radio con meccanismo di sincrotrone. La stella di neutroni centrale ha perso la sua importanza come sorgente di potenza per la nebulosa, benché possa ancora esserci un'emissione radio a impulsi. Il Nodo del Cigno è un esempio di residuo di supernova in questo stadio. Nello stadio finale (*5*) l'energia è stata completamente irraggiata e rimane un pulsar che può continuare, per milioni di anni, a emettere impulsi radio.

La produzione di interferon

Un RNA sintetico chiamato poli I:C può fornire una protezione a largo spettro contro le malattie virali. Esso stimola le cellule a produrre interferon, la proteina che difende altre cellule dall'infezione da virus

di Maurice R. Hilleman e Alfred A. Tytell

Potenzialmente vi sono tre metodi per controllare nell'uomo le malattie virali. Finora l'unico adottato è stato la somministrazione di un vaccino: una preparazione di virus uccisi o attenuati che stimola l'organismo a formare anticorpi contro quel determinato virus. La specificità dei vaccini, cioè il loro spettro ristretto, è una limitazione; ciò significa che è necessario un vaccino diverso per ciascun virus o ceppo di virus. Un altro metodo possibile, teoricamente dotato di azione a largo spettro, è la chemioterapia. Però non è stato scoperto nessun agente chimico non dannoso ed efficace che agisca contro i virus come i sulfamidici o gli antibiotici agiscono contro vasti gruppi di batteri. La terza possibilità è di affidarci a quella che apparentemente è la prima linea di difesa della cellula stessa contro l'attacco dei virus: l'interferon. Fin dal 1957 è noto che una cellula infettata da un virus produce una sostanza, l'interferon, che protegge altre cellule contro l'infezione virale. Ora è stato trovato un metodo efficace per mobilitare il sistema interferon dell'organismo ed è giunto il momento di valutarne con prove cliniche la pericolosità e l'efficacia.

Nel 1935 Meredith Hoskins della Rockefeller Foundation osservò che l'infezione di una scimmia con un virus della febbre gialla attenuato preveniva la morte dovuta a un virus della febbre gialla virulento somministrato contemporaneamente. Due anni dopo George M. Findlay e F.O. MacCallum della Wellcome Research Institution in Inghilterra hanno trovato che scimmie infettate con il virus della febbre della Rift Valley erano protette contro il virus della febbre gialla. Poiché questo effetto protettivo opera contro un secondo virus somministrato contemporaneamente a quello attenuato, e anche contro un virus differente, chiaramente non aveva nulla a che vedere con le note reazioni immunitarie che implicano la reazione antigene-anticorpo. Findlay e MacCallum l'hanno chiamata « interferenza virale », e negli anni seguenti sono stati riportati molti esempi di interferenza tra coppie di virus. Però nessuno trovò una via pratica per adattare il fenomeno alla prevenzione delle malattie virali.

Successivamente nel 1957, all'Istituto Nazionale per le Ricerche Mediche in Inghilterra, Alick Isaacs e Jean Lindenmann hanno fatto l'importante scoperta che l'azione di interferenza dei virus era espletata da una sostanza, che chiamarono interferon, che veniva prodotta dalle cellule infettate dal virus e che proteggeva le cellule non ancora infettate. L'interferon poteva essere distrutto con enzimi proteolitici, aveva un peso molecolare relativamente basso, ed era piú resistente agli acidi e alle basi della maggior parte delle proteine. Si è trovato che l'azione antivirale dell'interferon è molto vasta; questa proteina era attiva praticamente contro tutti i virus. D'altra parte aveva uno spettro ristretto in un altro senso: era attiva solo sulla stessa specie animale in cui era stata prodotta. L'interferon di topo era attivo solo in cellule di topo, l'interferon di pollo in cellule di pollo e cosí via.

La reazione immediata alla scoperta dell'interferon fu la speranza che la proteina potesse essere prodotta in cellule, purificata e somministrata per prevenire e curare le malattie virali. C'era anche la possibilità che la molecola di interferon contenesse un gruppo chimico, facilmente sintetizzabile, responsabile dell'attività antivirale di questa proteina. Il primo passo necessario era la purificazione, compito che fu intrapreso dal nostro gruppo al Merck Institute for Therapeutic Research. Nel 1963 potemmo annunciare un grado di purificazione significativo dell'interferon prodotto in embrioni di pollo che erano stati infettati col virus dell'influenza. L'interferon purificato era estremamente attivo contro i virus in esperimenti in vitro; esso era efficace come agente antivirale, per esempio, in quantità piú piccole di quelle necessarie ai comuni antibiotici contro batteri molto sensibili. Risultò che l'attività antivirale dell'interferon era dovuta alla complessa molecola nel suo insieme piuttosto che a una semplice porzione di essa. Le nostre osservazioni furono confermate e ampliate da Thomas C. Merigan jr. della Stanford University School of Medicine e da Karl H. Fantes dei Glaxo Laboratories Ltd., in Inghilterra.

È dubbio che sia stato finora isolato interferon assolutamente puro. Inoltre, non si è ancora trovato un metodo pratico per produrre interferon in cellule umane. Infatti, il rapporto tra dose richiesta e la possibile resa di interferon da parte di cellule è per ora cosí sfavorevole da potere escludere la possibilità di produrre interferon umano, purificandolo e somministrandolo agli uomini.

Considerando i limiti dell'uso di interferon esogeno in medicina, l'alternativa ovvia era di trovare un opportuno induttore di interferon: una sostanza che potesse essere data a un paziente per indurre il suo organismo a fabbricare il proprio interferon endogeno. La ricerca di induttori in molti laboratori portò alla scoperta di varie sostanze capaci di stimolare la produzione di interferon in animali. Queste includevano batteri, parassiti, virus, polisaccaridi, agenti come la fitoemoagglutinina che promuove la divisione cellulare, endotossine batteriche, plastiche sintetiche e altre sostanze. Per varie ragioni nessuno di questi agenti era adatto per uso clinico.

Il nostro gruppo al Merck Institute, costituito dagli autori e da A. Kirk Field, George P. Lampson e Marjorie M. Nemes, si è concentrato sul tentativo di trovare e poi imitare i meccanismi con cui i virus inducono naturalmente l'interferon. Decidemmo di seguire due vie. La prima era di frazionare un certo numero di virioni, o particelle virali, e provare a trovare uno o piú componenti responsabili dell'induzione dell'interferon. Il secondo era di

1 2 3 4 5 6 7 8 9 10
NUCLEO
VIRUS
CITOPLASMA
RNA
RNA REPLICATIVO
INTERFERON

Si pensa che il virus induca la produzione di interferon in due fasi successive. Il virus contenente RNA a filamento singolo si attacca alla parete cellulare (*1*) ed entra nella cellula (*2*). La capsula esterna del virus si distrugge liberando il filamento singolo di RNA (*3*) che si duplica e forma un filamento doppio (*4*). La presenza di un RNA esterno a doppio filamento fa sí che la cellula sintetizzi l'interferon (*5*) ma l'RNA continua a duplicarsi. L'interferon si trasferisce in una nuova cellula (*6*). Non può impedire a un virus di entrare in altre cellule (*7, 8, 9*), ma impedisce la duplicazione dell'RNA virale a filamento singolo (*10*).

Il poli I:C è un polinucleotide sintetico, analogo dell'RNA, realizzato in laboratorio unendo due omopolimeri: l'acido poliriboinisico (*a sinistra*) e l'acido poliribocitidilico (*a destra*). Ciascun filamento, come in un RNA naturale, è formato da uno scheletro (*in grigio*) di gruppi alternati di fosfato (*P*) e da ribosio a cui sono attaccate basi azotate come gruppi laterali. In questi omopolimeri sintetici le basi sono tutte uguali: la purina ipoxantina (*in colore*) e la pirimidina citosina (*in nero*). I due filamenti sono uniti attraverso le basi da legami di idrogeno (*linee tratteggiate*).

isolare il principio attivo della elenina, un estratto del fungo *Penicillum funiculosum* il quale, come Richard E. Shope e i suoi collaboratori al Rockefeller Institute for Medical Research avevano dimostrato, induce la produzione di interferon in colture di cellule e favorisce la resistenza all'infezione virale negli animali. Quello che ci mise sulla giusta strada fu la scoperta di Werber Braun e Masayasu Nakano della Rutgers University che la risposta anticorporale era innalzata da certi polinucleotidi sintetici. Questi sono analoghi degli acidi nucleici DNA e RNA che vengono prodotti in laboratorio combinando i nucleotidi (le subunità degli acidi nucleici) in maniera arbitraria anziché nella combinazione con cui si trovano in natura e che permette la codificazione dell'informazione genetica nelle cellule viventi.

Ci domandammo se questi polinucleotidi potessero indurre anche la formazione di interferon, e provammo vari RNA sintetici. Uno di questi risultò estremamente attivo: in piccole quantità (microgrammi) dava luogo alla produzione di interferon nei conigli e rendeva le colture di cellule e gli animali resistenti all'infezione virale. La sostanza attiva era il poli I:C, che consiste di filamenti omopolimerici accoppiati di acido poliriboinosinico (poli I) e acido poliribocitidilico (poli C) (*si veda la figura a sinistra*). Anche il poli A:U, in complesso basato sugli acidi adenilico e uridilico, era attivo sulle colture di cellule ma era molto meno attivo negli animali. Gli omopolimeri a filamento singolo non erano attivi.

Nel frattempo il lavoro sulla elenina procedeva bene; poco dopo aver stabilito la proprietà di indurre la produzione di interferon del poli I:C, potemmo affermare che anche il principio attivo dell'elenina era un RNA a doppia elica. Suggerimmo che la muffa da cui l'elenina veniva ottenuta doveva essere stata infettata da un virus, che aveva presumibilmente fornito l'RNA a doppia elica. I nostri sospetti furono confermati quando Walter J. Kleinschmidt e i suoi collaboratori dimostrarono la presenza di un virus dei funghi nello statalone, un estratto di un'altra specie di penicillo avente le stesse proprietà dell'elenina. Successivamente G.T. Banks e i suoi collaboratori dell'Imperial College of Science and Technology a Londra isolarono un virus dei funghi dal *P. funiculosum*.

Noi intraprendemmo successivamente un vasto programma di selezione provando vari tipi di polinucleotidi. Facemmo questo iniettando i polinucleotidi endovena in conigli, prelevando loro successivamente il sangue e saggiando la capacità del siero di proteggere una coltura di cellule contro l'infezione virale (*si veda la figura nella pagina a fronte*). Un gran numero di RNA a doppia elica, di origine naturale o sintetica, si dimostrarono attivi induttori. Gli RNA a filamento singolo erano inattivi, come anche il DNA a doppio filamento (*si veda la figura a pagina 28*).

Qual è l'origine dell'RNA a doppio filamento nel caso di una normale infezione virale? Mentre in alcuni virus, nei batteri, nelle piante e negli animali il materiale genetico è DNA, in molti virus è invece costituito da RNA. Nella maggior parte dei virus a RNA è la forma a elica singola di RNA a essere presente nel virione e non la forma a doppio filamento. Pensammo che tali virus quindi non potevano inizialmente dare luogo alla formazione di interferon; l'attività induttiva poteva insorgere solo quando nella cellula infetta il filamento di RNA veniva duplicato. È stato poi dimostrato mediante immunofluorescenza e altre tecniche che la forma a doppio filamento appare infatti nel normale corso della duplicazione virale. La stessa cosa vale per i polinucleotidi sintetici: gli omopolimeri diventano attivi solo quando sono appaiati a formare un doppio filamento. (In certe condizioni poli I o poli C da soli possono mostrare una debole capacità induttiva ma questa attività è trascurabile paragonata a quella del poli I:C e non ha quindi alcuna importanza pratica).

L'esatto meccanismo di induzione e utilizzazione dell'interferon deve ancora essere completamente chiarito, ma nella nostra ipotesi di lavoro riteniamo che il sistema di interferon operi in due fasi distinte (*si vedano le figure nelle precedenti due pagine*). Nella prima fase un virus contenente RNA a filamento singolo penetra nella cellula. Il suo rivestimento proteico esterno viene eliminato e si libera l'RNA che comincia a produrre il filamento complementare di RNA che è essenziale alla sua duplicazione. L'RNA a doppio filamento che si forma non è normale per le cellule, e così in qualche modo sconosciuto sollecita la cellula a sintetizzare interferon.

L'interferon di per sé non è attivo contro i virus. Lascia la cellula in cui è stato prodotto e, nella seconda fase, entra nelle cellule non infettate, dove esercita un'attività antivirale. Probabilmente funziona da « depressore » facendo produrre alla cellula una nuova proteina che impedisce la formazione di un acido nucleico virale nel caso che la cellula venga attaccata da un virus. Alternativamente, è possibile che l'interferon stesso sia alterato dalla nuova proteina, forse enzimaticamente, per formare una molecola antivirale attiva. Esiste soltanto una prova indiretta della presenza di questa nuova proteina e non è noto il suo preciso meccanismo d'azione nel prevenire la duplicazione virale. I pochi virus (come i reovirus) che contengono RNA a doppio filamento inducono l'interferon immediatamente al momento della penetrazione della cellula senza bisogno di duplicazione.

Per valutare l'attività protettiva su colture di cellule, l'induttore in esame viene aggiunto alla coltura e circa tre ore dopo viene somministrata una dose stabilita e infettiva di virus. L'attività è dimostrata dalla capacità che la sostanza ha di prevenire il danno da virus nello strato di cellule in coltura. Tutti gli RNA a doppio filamento trovati attivi, nell'indurre la produzione di interferon nei conigli erano anche attivi, in quantità di microgrammi, nel prevenire la morte cellulare dovuta al virus della stomatite vescicolare in un'ampia varietà di cellule di uccelli e mammiferi, compreso l'uomo. Quando gli stessi induttori venivano provati sui topi prevenivano la malattia e la morte in animali infettati con il virus della polmonite del topo.

Per la facile disponibilità di poli I:C e per i suoi promettenti risultati in questi primi esami, gli studi successivi nei nostri laboratori furono rivolti soltanto a questa sostanza. Seguendo una comune pratica per dosare l'efficacia di un medicamento, abbiamo scelto come « sistema modello » i topi, infettandoli con una dose letale standard di virus diversi in modo da poter misurare l'effetto profilattico e terapeutico di poli I:C. In un esperimento tipico questa sostanza veniva instillata nelle narici del topo prima o dopo l'inoculazione nasale di una quantità letale standardizzata di virus della polmonite (*si veda la figura a pagina 29*). Quasi tutti i topi di controllo non trattati morivano, mentre tutti quelli trattati con poli I:C prima del virus sopravvivevano. Lo stesso avveniva per quasi tutti i topi a cui fu dato poli I:C due giorni dopo il virus. In altre parole, il poli I:C aveva un effetto sia terapeutico che profilattico. Evidentemente il poli I:C, dato abbastanza presto dopo l'infezione, fa produrre interferon sufficiente a proteggere una grande percentuale delle cellule; in effetti l'interferon ha la meglio sul virus e l'animale sopravvive. Se il trattamento viene iniziato quattro giorni dopo l'inoculazione del virus si rivela inefficace. Esperimenti di questo genere hanno rivelato un forte effetto protettivo del poli I:C nei topi contro il virus dell'encefalomiocardite, il virus vaccinico e i virus parainfluenzali. Effetti protettivi minori sono stati ottenuti negli animali contro il virus oncogeno (che produce cancro) SV-40, l'adenovirus e il virus della leucemia di Friend. Questa capacità di proteggere contro virus contenenti sia DNA che RNA, contro virus oncogeni e contro

Una prova per dimostrare la presenza di interferon si fa iniettando nei conigli poli I:C (o qualsiasi altra sostanza potenzialmente induttrice) per via endovenosa. Dopo due ore si preleva il sangue dai conigli e si lascia coagulare. Il siero viene quindi esaminato a varie diluizioni su cellule in coltura (*a*); la coltura viene incubata per 18 ore, lavata e poi infettata con il virus della stomatite vescicolare. Se il livello dell'interferon è quello ottimale, le cellule non sono danneggiate. Contemporaneamente una coltura di controllo non trattata viene esposta al virus allo stesso modo (*b*). Le sue cellule vengono uccise dal virus.

MATERIALE	DOSE (MICROGRAMMI)	TITOLO DELL'INTERFERON
RNA A DOPPIO FILAMENTO		
POLI I:C	2	640
POLI A:U	25	20
PENICILLIUM FUNICULOSUM	8	640
REOVIRUS 3	8	640
COLIFAGO MS 2 (FORMA REPLICATIVA)	8	160
COLIFAGO MUTANTE MU 9 (FORMA REPLICATIVA)	2	40
VIRUS DEL RISO NANO	20	640
VIRUS DELLA POLIEDROSI CITOPLASMICA	22	160
RNA A FILAMENTO SINGOLO		
POLI I	25	0
POLI C	20	0
COLIFAGO MS2	8	0
ESCHERICHIA COLI	100	0
VIRUS DELLA MALATTIA DI NEWCASTLE	10	0
VIRUS DELL'INFLUENZA	10	0
VIRUS DEL MOSAICO DEL TABACCO	40	0
RIBOSOMI DI LIEVITO	1000	0
RNA SOLUBILE DI LIEVITO	200	0
DNA A DOPPIA ELICA		
TESSUTO DI TIMO DI VITELLO	200	0

La tabella indica la capacità di vari acidi nucleici nell'indurre la produzione di interferon. Gli RNA a doppio filamento sono efficaci a dosi di microgrammi. A destra sono indicate le diluizioni di siero di coniglio attive nel proteggere il 50 per cento delle colture di cellule. RNA a filamento singolo e DNA a doppio filamento sono inattivi.

quelli che uccidono le cellule, sottolineava l'attività antivirale a largo spettro del poli I:C.

Altri ricercatori hanno esteso ampiamente gli studi su poli I:C. Essi hanno dimostrato l'efficienza del poli I:C contro gravi infezioni di *Herpes simplex* nell'occhio del coniglio, contro infezioni del virus Semliki Forest nel topo (un virus portato dagli insetti), contro la rabbia nei conigli e contro certe infezioni di virus respiratori.

Inoltre sono stati osservati per il poli I:C notevoli effetti antitumorali contro parecchi tumori trapiantati nel topo, in particolare il tumore dell'ascite L 1210. È stata anche riscontrata la sua attività contro la malaria sperimentale e altre malattie parassitarie.

Come si è notato sopra si sa che poli I:C stimola i comuni meccanismi immunitari, compresa l'immunità cellulare, e si pensa che questo sia importante nella difesa contro i tumori.

Prima che una medicina possa essere utilizzata in esperimenti clinici sull'uomo, bisogna assicurarsi che non sia dannosa mediante esperimenti su animali; abbiamo così fatto prove su topi, ratti, cani, scimmie con differenti dosi e diverse vie di somministrazione. Quando questa medicina veniva iniettata endovena, si notavano effetti collaterali a livello dei capillari sanguigni, del fegato e degli organi ematopoietici dei cani; questi effetti erano minori o assenti con somministrazione per via sottocutanea o respiratoria. Gli effetti tossici erano molto minori nei topi e nei ratti che nei cani, ed erano essenzialmente assenti nelle scimmie. Questi esami indicavano che ci sarebbe un soddisfacente margine di sicurezza per soggetti umani trattati con dosi di poli I:C abbastanza elevate da stimolare l'interferon.

Altri controlli sulla sicurezza hanno dimostrato che il poli I:C non provoca nella cavia reazioni anafilattiche. Questa sostanza non era cancerogena se iniettata in animali neonati che furono osservati fino a 18 mesi dopo l'iniezione. Non provocava l'apparire di nessun fenomeno di tipo canceroso in colture di cellule umane normali anche se venivano impiantate nella borsa guanciale del criceto. E infine non provocava alcune alterazioni nei cromosomi di cellule umane normali coltivate in vitro.

Quando venne il momento di fare prove sull'uomo, considerando il notevole effetto del poli I:C contro certi virus che provocano il cancro negli animali e l'azione sorprendentemente benefica di questa sostanza contro certi tumori trapiantati, decidemmo di provarla su malati di cancro. Per questo facemmo uno studio in collaborazione con Charles W. Young e Irwing H. Krokoff del Memorial Hospital e lo Sloan Kettering Institute for Cancer Research a New York. Venti pazienti affetti da cancro avanzato di tipi diversi vennero trattati col poli I:C per via endovena, da due microgrammi a quattro milligrammi per chilogrammo di peso corporeo. Vennero presi campioni di siero per dosare l'interferon da prima dell'iniezione fino a 72 ore dopo. Quattordici pazienti su venti reagiscono producendo interferon.

L'induzione della produzione di interferon e il grado e la durata della risposta non sembravano essere molto in relazione con la dose del poli I:C. L'interferon compariva nel sangue due ore dopo l'iniezione di poli I:C e persisteva in alcuni pazienti fino a 72 ore. La massima produzione di interferon veniva generalmente raggiunta tra 12 e 48 ore dopo la somministrazione della sostanza. Sfortunatamente in questi pazienti non c'era nessun effetto benefico significativo contro i tumori.

Due problemi fondamentali che riguardano l'applicazione clinica delle sostanze che stimolano la produzione di interferon sono la durata relativamente corta della resistenza dopo somministrazione dell'induttore (in genere circa sei giorni negli animali) e la mancanza di risposta a una seconda iniezione dell'induttore in un periodo immediatamente successivo. Per fortuna questo stato di iporeattività sembra scomparire quasi contemporaneamente alla perdita di resistenza all'infezione virale, proprio quando c'è bisogno di dare una nuova dose di induttore. Negli studi fatti su pazienti si sono ottenute riinduzioni positive di interferon in due casi quando iniezioni di poli I:C venivano somministrate alla distanza di sei o sette giorni e in un paziente persino di tre giorni. Un altro paziente cominciò a non reagire solo dopo la quarta iniezione di una serie di dosi giornaliere di poli I:C, e questo paziente diventò di nuovo completamente sensibile dopo un periodo di 10 giorni. Sembra quindi che, sebbene la resistenza alle sostanze che inducono la produzione di interferon si verifichi nell'uomo, la sua durata sia relativamente breve. Può essere che la somministrazione di piccole, ma terapeuticamente adeguate dosi di poli I:C a intervalli frequenti permettano una produzione continua di interferon mantenendo così una continua resistenza alla infezione virale.

Nell'uomo le reazioni cliniche al poli I:C erano scarse. La reazione più notevole a iniezioni endovenose di poli I:C era la febbre. L'innalzamento di temperatura poteva arrivare a circa tre gradi raggiungendo il massimo da sei a quindici ore dopo l'iniezione. La febbre era in genere accompagnata dalla produzione di interferon, sebbene in alcuni pazienti le due reazioni si verificavano indipendentemente l'una dall'altra. L'aumento della temperatura non era correlato con la dose di poli I:C. Ampie analisi di laboratorio non hanno rivelato disturbi alle funzioni del fegato, del rene o del midollo osseo. Non fu osservato alcun effetto sul meccanismo della coagulazione del sangue. Insomma, non apparivano sintomi clinici che ponessero un limite alla dose somministrata.

Ci interessava anche la possibilità di una risposta immunitaria al poli I:C, ma le analisi hanno dimostrato che i pazienti non formavano anticorpi né contro il poli I:C, né contro il DNA. C'è ancora un'ulteriore possibilità di investigare. Si è notato che lo svilupparsi di una malattia autoimmunitaria che ricorda un diffuso eritema nell'uomo, e che appare spontaneamente in una particolare razza di topi, può essere accelerato mediante iniezioni multiple di poli I:C. L'entità della dose e la via di somministrazione di poli I:C in questi topi vengono attualmente studiate per stabilire un altro limite di sicurezza per l'uso di questa sostanza nei soggetti umani suscettibili a malattie immunitarie.

La maggior parte delle infezioni virali nell'uomo sono auto-limitate suggerendo che i meccanismi naturali di difesa sono attivi durante la guarigione da queste malattie. Gli anticorpi e l'immunità cellulare, i comuni meccanismi immunitari specifici, sembrano entrare in azione in una fase piuttosto tarda del processo infettivo e possono avere maggiore importanza nel prevenire una successiva reinfezione da parte dello stesso virus o nell'eliminare cellule infettate dal virus piuttosto che nell'affrettare la guarigione dell'ospite da una particolare infezione virale. L'interferon viene prodotto nelle prime fasi del processo infettivo e ci sono buoni motivi per credere che esso sia per lo meno uno degli importanti fattori implicati nella guarigione da una malattia virale. La somministrazione di interferon esogeno ha dato fino a ora risultati poco promettenti a causa delle difficoltà di procurarlo, della sicurezza e del costo. La via più pratica sembra essere la somministrazione di sostanze capaci di indurre la produzione di interferon. Il fatto che possano verificarsi effetti secondari benefici, come la stimolazione degli ordinari meccanismi immunitari, è una ragione in più per utilizzare i polinucleotidi sintetici. La bassa tossicità per l'uomo e la pronta disponibilità di questo prodotto schiude la possibilità di ricerche su vasta scala. Dopo alcune prove definitive per eliminare il pericolo di disturbi autoimmunitari, questo prodotto sarà pronto per caute prove su soggetti umani per prevenire parecchie infezioni, come il comune raffreddore, che sono causate da diversi virus.

1 POLI I:C — PLACEBO

TRE ORE

2 VIRUS

14 GIORNI

3

Per analizzarne l'effetto protettivo, poli I:C venne instillato nelle narici di 17 topi (*a sinistra*) e a 32 controlli (*a destra*) fu dato un placebo. Dopo tre ore ambedue i gruppi vennero infettati con una dose letale di virus della polmonite. Gli animali vennero poi osservati per 14 giorni. Tutti i topi trattati sopravvissero, mentre sopravvisse soltanto un controllo. La sostanza era efficace anche quando veniva somministrata due giorni dopo il virus; la sopravvivenza era di 15 su 15 tra gli animali trattati e 1 su 25 tra i controlli. Poli I:C non aveva effetto se dato quattro giorni dopo il virus.

La psicoterapia della famiglia

Dal 1967 opera a Milano il primo centro italiano di psicoterapia della famiglia. In questo articolo se ne illustrano le premesse teoriche fondamentali e i risultati clinici sinora conseguiti

di Mara Selvini Palazzoli

Da poco più di dieci anni, la psichiatria della famiglia ha potuto evolvere da una posizione relativamente oscura a un riconoscimento sempre più vasto sulla scena delle scienze umane.

Come era già ripetutamente avvenuto per altri movimenti evolutivi in psichiatria, e segnatamente per la psicoanalisi, il « movimento » per la famiglia trovò la sua spinta primigenia non in un interesse concettuale, teoretico, ma in una motivazione clinica: lo sforzo di trovare metodi di cura più adeguati per trattare e risolvere specifici problemi emozionali, particolarmente quelli che sembravano sfuggire, o resistere, alle metodiche terapeutiche individuali. In certo senso, il « movimento » della famiglia si sviluppò dapprima come una estensione della psicoanalisi, dopo che questa aveva vinto la lunga e dura battaglia del periodo iniziale, e, a partire dagli anni trenta, era stata generalmente accettata, non solo come strumento per aiutare singolarmente i sofferenti di disturbi psicoemotivi, ma anche come modello concettuale per la comprensione dell'uomo. È infatti nei centri di lavoro psicoanalitico che apparvero, dapprima negli Stati Uniti, i primi segni del graduale orientamento verso la famiglia come oggetto, sia pure ancora marginale, di interesse psichiatrico, precisamente nei centri psicopedagogici e nelle cliniche dedicate al recupero di bambini socialmente disadattati e psichicamente disturbati.

Oltre alla applicazione e allo sviluppo delle tecniche di psicoanalisi infantile, nel programma di lavoro dei centri sopra ricordati divenne procedura abituale che un primo terapista si occupasse individualmente del bambino, mentre un secondo terapista (spesso un assistente sociale opportunamente preparato) lavorava con i genitori. Si considerava tuttavia primario il trattamento individuale del bambino, continuando a considerare sussidiario il lavoro condotto separatamente con gli adulti della sua famiglia. Infatti la teoria psicoanalitica era stata formulata sulla base dello studio dettagliato del paziente individuale. I dati sulla famiglia venivano derivati assai più dalle percezioni «personali» del paziente che non dall'osservazione diretta. In base a questa posizione teorica, il fulcro della attenzione era centrato sul paziente, mentre la famiglia restava al di fuori dell'immediato interesse teoretico e terapeutico. La teoria psichiatrica individuale era ancora basata sul modello medico, coi suoi concetti tradizionali di eziologia, diagnosi, patologia, classificazione nosografica e cura della « malattia » come fenomeno individuale.

Il paziente era considerato vittima impotente di un « qualcosa » che era insorto dentro di lui e che si sottraeva al suo controllo.

Il dilemma concettuale iniziò a porsi allorché le persone più importanti e significative nell'esistenza del bambino cominciarono a venir considerate « la causa » dei suoi comportamenti anormali, e perciò patogenetiche.

Insoddisfatti, e a ragione, di tale concezione, si tentò in vari modi, di mitigarne la durezza. Facendo ricorso all'inconscio, si postulò che i genitori potevano inconsapevolmente danneggiare il bambino forse anche nel tentativo di aiutarlo. Ciò era già diverso dal giudicare che il danno fosse arrecato intenzionalmente, o per grave carenza di senso di responsabilità: tuttavia i genitori restavano « patogenici », erano cioè considerati, se non concettualmente, almeno emotivamente, causanti la patologia dei figli.

Il passo decisivo fu compiuto negli ultimi dieci anni quando la cornice che isolava l'individuo venne infranta, e alle domande del perché un uomo fa quello che fa si rispose non più cercando « qualcosa » dentro di lui, ma riferendosi al contesto di rapporti che egli concrea e nei quali vive. Tale contesto di interscambi è stato indicato con la designazione di *sistema* in quanto esso si costituisce inevitabilmente ogniqualvolta un gruppo di persone si trova in intima e durevole comunanza di vita e presenta caratteristici legami integrativi che seguono determinati modelli di funzionamento.

Il fatto di aver posto l'accento sui rapporti interpersonali e i recenti sviluppi delle idee relative ai sistemi era alla base di questo cambiamento. La modificazione fondamentale è costituita dallo spostare l'attenzione dell'individuo e la sua natura ai *modelli abituali e sistematici di comportamento* che gli uomini elaborano nei rapporti con gli intimi. Da qui si ha l'estensione delle ricerche sulla famiglia e della terapia della famiglia.

« In questi ultimi anni per la prima volta nella storia, coppie sposate – ha scritto J. Haley – e famiglie intere vengono sottoposte a osservazione sistematica. I membri delle famiglie vengono osservati mentre interagiscono. Il problema per la ricerca è come concettualizzare il comportamento-risposta ripetuto in questo tessuto sociale attivo in modo che permangano vere nel tempo le affermazioni sulle regolarità nell'interscambio. Il problema non è più di caratterizzare e classificare questi individui: ma di descrivere e classificare i modelli *abituali* di comportamento-risposta che vengono scambiati tra persone in stretto rapporto. È possibile categorizzare i processi tipici del gruppo, descrivere le modificazioni che eventualmente si producono e differenziare un'organizzazione dall'altra? Con questa impostazione, la "causa" delle azioni di una persona si sposta da ciò che è dentro di lei al contesto nel quale essa vive. Quando ci si chiede se una famiglia che ha tra i suoi membri un paziente psichiatrico (paziente designato) costituisca un tipo di organizzazione diverso da un'altra famiglia che non lo presenta, equivale a chiedersi se esista un "sistema di comportamento interattivo" che provochi o richieda in uno o più dei membri del sistema un comportamento classificabile come psicopatologico. Questi interrogativi comportano dei riflessi che vanno al di là del settore delle ricerche sulla famiglia. Nella misura in cui ogni gruppo di uomini ha una storia e uno sviluppo comune (per esempio i gruppi di ricerca o i gruppi di aziende commerciali) il tipo di scambi generati nel sistema che tali gruppi costituiscono e con cui funzionano può determinare il fatto che vi siano dei rapporti amichevoli e un lavoro produttivo, o che vi sia invece un gruppo di partecipanti infelice e disturbato. Nel più vasto ambito sociale, la famiglia delle nazioni può elaborare modelli di scambio prevedibilmente generatori di rovina e di guerra ». Per analizzare questi interrogativi occorre spostare l'attenzione dallo studio dell'individuo allo studio dei rapporti fra le parti (sottosistemi) interagenti in un più ampio sistema. Ciò può essere meglio espresso con la classica formulazione transazionale che gli americani Spiegel e Bell pubblicarono nel 1959, attualmente alla base della più avanzata ricerca psicosociale. Entro il campo totale che racchiude i sottosistemi tra loro connessi, un sottosistema componente, quale l'individuo, potrebbe essere isolato e studiato come una entità, ma questo è un artificio euristico che comporta sempre qualche deformazione e o qualche perdita di precisione e di capacità di previsione. Se questa posizione è accettata, allora l'individuo e la famiglia sono visti come sottosistemi interrelati in modo intimo e complesso. Nessuno dei due ha alcuna « realtà » maggiore dell'altro, nessuno dei due ha diritto ad alcuna priorità o causalità. Ciò sposta il problema da « che cosa nella famiglia *causa* la patologia nell'individuo? » a « quali processi che hanno luogo tra l'individuo e la famiglia sono collegati a quei comportamenti che chiamiamo patologici? ». Questo significa provvedersi di un'ottica interamente nuova per indagare sul funzionamento di sistemi complessi di interazione psicosociale. Per ritornare allo studio della famiglia, il salto concettuale consistette nel fissare l'attenzione sulla famiglia come unità. Si risolse cioè di abbandonare lo studio della personalità individuale entro la sua famiglia per volgersi a un più complesso punto di vista: la interdipendenza dei membri della famiglia come membri di un sistema. Nel campo specifico della ricerca psichiatrica, il cosiddetto paziente non è più il membro ammalato della sua famiglia, ma piuttosto l'espressione, o meglio il sintomo della disfunzione del gruppo familiare considerato come una unità di cui il paziente stesso è membro concreante. Il risultato di questa nuova concezione è il considerare la famiglia non più come agente nocivo, patogenico, ma come unità sofferente bisognosa di aiuto.

Gli strumenti utilizzati per la ricerca sulla famiglia, possono classificarsi in tre gruppi:

1) interviste o resoconti ottenuti dagli stessi membri della famiglia mediante questionari;

2) osservazione dei membri della famiglia mentre interagiscono insieme, in situazioni sperimentali, strutturate in modo il più possibile standardizzato

La psicoterapia familiare viene praticata in sedute alle quali partecipa tutto il gruppo familiare in trattamento. È visibile nella illustrazione lo specchio unidirezionale oltre il quale è la camera di osservazione per la registrazione delle sedute.

(esecuzione di compiti familiari prefissati, reattivi psicodiagnostici somministrati collettivamente, ecc.). L'osservazione diretta è corredata da registrazioni su nastro, registrazioni video e filmati;

3) trattamento psicoterapico del gruppo familiare, cioè del gruppo familiare in toto, ivi compreso il paziente designato. Tale trattamento psicoterapico viene condotto dalle varie scuole o mediante un solo terapeuta, o con due coterapeuti, oppure, come nel Centro di Houston, con l'intervento contemporaneo di piú coterapeuti. Anche nelle sedute psicoterapiche si fa regolarmente uso di registrazioni su nastro o di registrazioni video. I trascritti delle registrazioni vengono poi sottoposti ad analisi minuziosa da parte dell'équipe dei ricercatori.

Questi diversi strumenti di studio della famiglia ovviamente raccoglieranno dati diversi.

Tuttavia, lo strumento della ricerca che ha offerto sinora una piú larga messe di dati sembra essere la psicoterapia familiare. E ciò in totale accordo con la nota affermazione di Kurt Lewin, iniziatore della terapia di gruppo, il quale sosteneva: « se vuoi davvero vedere come una cosa funziona, tenta di cambiare il suo funzionamento ». In senso lato, i terapisti divennero degli esperti nel capire i sistemi familiari disfunzionanti e nell'aiutare la famiglia a modificarli per costruire un equilibrio funzionale non piú rigido, ma aperto elasticamente al progresso e alla modificazione.

Il movimento per lo studio e la psicoterapia della famiglia, sinora è stato poco sviluppato in Europa (con l'eccezione dell'eccellente scuola dello psichiatra Alaneh, in Finlandia, che da anni conduce un'importante ricerca sistematica sulle famiglie con pazienti schizofrenici). Negli Stati Uniti, invece, come già si è accennato, ha al suo attivo una storia ormai pressoché trentennale con una serie di scuole e di istituti di ricerca con indirizzi concettuali, teorici e sperimentali e una bibliografia ormai vastissima, rappresentata in centinaia di saggi e in decine di volumi.

Allo scopo di agevolare il compito non facile di chi voglia approfondire la letteratura sulla psicoterapia della famiglia, Beels e Ferber, in una recente revisione critica, hanno raggruppato i terapeuti della famiglia statunitensi in tre gruppi, designando ciascuno di questi con un termine che ne caratterizza sinteticamente lo spirito e lo stile di lavoro: i « conductors » o conduttori, i « system purists » o puristi del sistema, e i « reactors analysts », disegnazione difficilmente traducibile, ma comunque corrispondente nel significato a psicoanalisti reattivi.

I *conductors*, (Ackerman, Satir, Bowen, Minuchin) lavorano singolarmente, in modo rapido e attivo, alla modificazione del sistema di funzionamento familiare, specie in casi di crisi acuta, ponendo se stessi al centro della situazione come catalizzatori di una comunicazione aperta e senza ipocrisie. Essi usano precipuamente di se stessi, della forza della propria personalità, e sono anche convinti assertori della validità della propria Weltanschaung e della propria scala di valori. (Posizione, quest'ultima, che non condividiamo, in accordo con molti altri terapeuti della famiglia, in quanto rispecchia una mentalità costrittiva e in sostanza morfostatica). Ackerman è veramente un conduttore tipico. Egli mobilizza l'interazione familiare, attento a non lasciarsi sfuggire gesti e comportamenti non verbali, indizi anche minimi di comunicazioni interpersonali. Taglia corto coi tentativi piú o meno scoperti di negazione, ipocrisia e proiezione, forzando i membri della famiglia a essere piú sinceri e aperti con lui di quanto non lo siano tra di loro. Con il suo atteggiamento confidenziale e diretto egli li impegna subito in una relazione demistificante. Disgela una famiglia congelata dal terrore dell'aggressività, (caso quanto mai frequente anche nella nostra esperienza) o promuove lo scambio sessuale in una coppia che ha paura del sesso. Rompe le « regole » segrete della famiglia, quelle regole che concernono ciò che non si deve mai nominare, giacché non esistono per lui cose innominabili, e insegue ciò che la gente tenta di nascondere finché riesce ad afferrarlo. Fatto questo, diventa mallevadore, impresta alla famiglia il suo piacere di vivere, il suo humor, il suo apprezzamento di una soddisfacente vita sessuale e una aggressività aperta, ma controllata.

I *system purists*, in maggioranza della scuola di Palo Alto (Jackson, Haley, Zuk Watzlawick) s'occupano esclusivamente del sistema di funzionamento familiare, cioè della modalità comunicativa, verbale e non verbale, dei partecipanti, specie per quanto riguarda gli aspetti prammatici della comunicazione. Gli aspetti prammatici della comunicazione indicano quelle componenti del messaggio che provocano un determinato effetto sul comportamento dell'interlocutore, esattamente come la mossa del giocatore di scacchi influenza inevitabilmente la successiva mossa dell'avversario. Benché la spiegazione storica della posizione dei pezzi, a un dato momento del gioco, possa essere ottenuta ricostruendo le mosse fatte fin dall'inizio dei giocatori, i problemi cruciali al momento dell'osservazione sono: qual è il rapporto dei pezzi fra loro? quali sono le regole che governano i giocatori? quale sarà la prossima mossa? Il terapista è la persona a cui i membri della famiglia si presentano con un sistema di comunicazione « inchiodato » indefinitivamente in un circolo chiuso. Egli deve intervenire per chiarire e cambiare le regole che imprigionano inconsapevolmente i partecipanti.

I *reactors analysts* (Wynne, Whitaker, B. Nagy, Framo) si distinguono per una terminologia e un interesse piú prossimi alla tradizione della terapia psicoanalitica, e sono, come indica il termine stesso, piú reattivi che attivi nella condotta terapeutica. Pur utilizzando largamente la chiarificazione delle modalità comunicative, verbali e non verbali, essi tengono presenti, anche se li interpretano raramente o tardivamente, i moventi dinamici profondi dei singoli e del gruppo. Tuttavia, nonostante la patente differenza degli indirizzi concettuali, ideologici, e tecnici tra le varie scuole, ciò su cui tutti i terapisti della famiglia possono essere d'accordo è lo scopo finale del loro lavoro: « rendere la famiglia un gruppo ben funzionante ». Benché differenti terapisti potranno divergere su come definire un buon funzionamento, e qual è il rapporto di questo con lo stato degli individui che lo compongono, i terapisti sono in genere d'accordo che lo scopo di lavorare con un gruppo familiare è di promuoverne « il funzionamento come famiglia ». Tale definizione è particolarmente inclusiva se si tiene presente che i terapisti della famiglia considerano, tra le funzioni piú importanti della stessa, la promozione della differenziazione e individuazione dei suoi membri come persone autonome e distinte. Il terapista della famiglia non si sforza primariamente di rendere la famiglia piú « gruppale », piú coesa, ma si sforza innanzitutto di promuovere nei componenti le singole potenzialità di crescita e di differenziazione. In sostanza, si tratta di promuovere simultaneamente ciò che potrebbe a prima vista apparire antinomico: la integrazione funzionale nel gruppo e contemporaneamente la differenziazione dei singoli.

Il movimento per lo studio della famiglia in Italia, ha inizi abbastanza recenti. Solo da circa quattro anni, anche in seguito a una certa azione di pressione sull'ambiente psichiatrico esercitata dalla pubblicazione di alcune

opere straniere sull'argomento, l'idea incomincia a farsi strada, e alcune équipes per la psicoterapia e lo studio della famiglia incominciano a costituirsi, per lo piú nell'ambito di istituzioni psichiatriche.

Io mi limiterò qui, tuttavia, a parlare del Centro di ricerca in cui lavoro da alcuni anni, e che si è costituito a Milano sotto la denominazione di « Centro per lo studio della famiglia e delle tecniche di gruppo ». Mi pare interessante premettere alcune notizie sulla storia e sugli inizi della attività del Centro stesso, specie per quanto si può derivarne dal punto di vista socioculturale.

Il Centro ha iniziato la sua attività a Milano nel maggio 1967. L'apertura del lavoro era stata preceduta da un lungo periodo di studio e di approfondimento teoretico. Nel marzo 1967, dopo aver preso parte alla « Conference on systematic research on family interaction » svoltasi il 17 e 18 marzo presso l'Estern Pennsylvania Psychiatric Institute di Filadelfia, con la partecipazione dei piú noti esperti dell'argomento, fui ospite di quell'Istituto per un breve periodo, allo scopo di osservare lo svolgimento dell'attività di corso, e di partecipare alle discussioni di équipe.

Sistemata l'attrezzatura indispensabile – una sala spaziosa arredata per le riunioni familiari, una camera di osservazione con specchio unidirezionale (*si vedano le illustrazioni alle pagine 31 e 34*), nonché l'apparecchiatura per la sonorizzazione e la registrazione su nastro – iniziammo l'attività clinica e di ricerca.

Questa si svolse per circa un anno e mezzo fra molte difficoltà di carattere pratico e in misura ridotta, dovuta sia alla difficoltà di reperire o motivare le famiglie al trattamento in un ambiente quasi totalmente impreparato o decisamente ostile a tale orientamento, sia per l'esiguità dell'équipe, formata allora da due soli coterapeuti. Nel 1969 l'équipe è salita a 5 terapeuti, 3 uomini e 2 donne, tutti medici psichiatri, con training compiuti per la psicoterapia individuale e di gruppo. Recentemente è entrato a far parte dell'équipe anche uno psicologo che si dedica precipuamente all'ordinamento e all'elaborazione teoretica dei dati ottenuti nel corso del trattamento.

Per vari motivi, su cui non è qui il caso di dilungarsi, ma certamente in rapporto con l'attuale situazione della psichiatria in Italia, si è preferito mantenere il Centro totalmente indipendente da pubbliche istituzioni, quali i Consultori di igiene mentale, le Cliniche psichiatriche universitarie, gli Ospedali psichiatrici. Sono infatti noti i fenomeni di pressione che possono essere esercitati su di una équipe amministrativamente non autonoma, quali fissazioni di scadenze per la pubblicazione di lavori, imposizioni dall'esterno di nuovi componenti dell'équipe, strumentalizzazione dell'esperto a scopi politici.

Tale decisione di autonomia ci pose tuttavia di fronte ad altre difficoltà, tra cui la mancanza, per il momento, di qualsiasi sovvenzione per la ricerca. Abbiamo aggirato l'ostacolo lavorando part-time, con un numero ridotto di famiglie che in maggioranza sostengono un rimborso spese che viene stabilito proporzionalmente ai loro mezzi economici. Fondata legalmente una « Associazione per lo studio della famiglia » devolviamo gli introiti alle spese di mantenimento dell'istituto e della ri-

Rappresentazione grafica del sistema di funzionamento di una coppia secondo lo schema che Paul Watzlawick ha mutuato dal concetto matematico delle serie infinitamente oscillanti di Bolzano. La moglie è invadente e brontolona, il marito chiuso e taciturno. La moglie dirà di comportarsi cosí « a causa » dell'atteggiamento del marito; il marito dirà di comportarsi a quel modo « a causa » dell'atteggiamento della moglie. Ma in questo caso l'interpunzione che ciascuno dei due pone nella serie delle sequenze comunicative allo scopo di definire un principio causale, è arbitraria. Il sistema di funzionamento che noi osserviamo è complementare, simultaneo, sottratto al principio di causalità e come tale è l'unica spiegazione di se stesso.

Camera d'osservazione del Centro di psicoterapia della famiglia di Milano. Altri membri dell'équipe possono, attraverso lo specchio unidirezionale, osservare i comportamenti dei vari membri del gruppo, compresi i terapeuti, portando poi nella discussione di équipe che segue a ogni seduta il contributo, spesse volte prezioso e di obiettiva utilità, di osservatori imparziali.

cerca. Tali introiti sono costituiti sia dalle quote degli associati che dai contributi delle famiglie. I nostri pazienti, o piuttosto le nostre famiglie, appartengono in maggioranza alla classe media superiore e a quella delle professioni liberali, in piccolo numero alla classe media inferiore (impiegati, artigiani, operai qualificati). È interessante notare a questo punto come, nonostante che l'aggiornamento culturale e il livello di informazione su tale tipo di trattamento sia indubbiamente piú elevato nelle classi superiori, sia assai piú facile motivare alla terapia familiare, ottenendo rapidamente un rapporto di fiducia e collaborazione, le famiglie dei ceti meno abbienti. Le « grandi famiglie » (aristocratici, borghesi di alta tradizione) legate a complessi pregiudizi di clan, hanno sinora impavidamente declinato ogni tentativo di offerta terapeutica, preferendo affidare i pazienti designati (talvolta piú di uno per famiglia) alle cure di analisti individuali o di cliniche straniere private. Ci sembrano, questi, fenomeni di grande interesse, sul cui approfondimento intendiamo insistere nell'avvenire. Le famiglie al completo sono talvolta inviate da medici psicoterapisti. Piú spesso, tuttavia, l'iter terapeutico inizia con la richiesta di una terapia individuale del paziente designato a qualche terapeuta della nostra équipe. Se questi constata un forte coinvolgimento dei problemi del paziente con quelli della famiglia, cioè difficoltà di rapporto che insorgono nello schema di interazione reciproca della coppia o della famiglia, a cui tutti, collusivamente o apertamente contribuiscono, propone alcune sedute di gruppo familiare. Tali sedute hanno luogo al Centro, in vista di una eventuale decisione terapeutica.

Giacché le famiglie che volontariamente acconsentono alla psicoterapia debbono anche sostenere una spesa, si può dedurre che le loro motivazioni sono paragonabili a quelle dei pazienti che volontariamente entrano in una psicoanalisi individuale. Il sostenere una spesa presuppone una attitudine di collaborazione attiva, salvaguardando inoltre l'utilissima attitudine critica delle famiglie nei riguardi dei terapeuti e la libertà del rapporto. Proprio ciò che fa difetto, non raramente, nei trattamenti psicoterapici di famiglie non paganti, le cui spese sono sostenute da fondi pubblici, a scopo sociale o di esperimento. Nella nostra esperienza, che ebbe, specie nei primi due anni, momenti molto frustranti, la libertà del rapporto connessa al pagamento fu la spia piú sensibile dei nostri errori terapeutici, che ci venivano puntualmente dimostrati e con la critica aperta e, piú spesso, con il troncamento della terapia.

La nostra organizzazione, tuttavia, pure se ideale per strutturare un autentico contesto psicoterapico, presenta al momento indubbi svantaggi per ciò che riguarda la ricerca: ci limita a una casistica relativamente ristretta, cioè a notevoli difficoltà di reperimento, qualora si desideri approfondire la ricerca su problemi determinati (per esempio, su famiglie con pazienti affette da anoressia mentale, allo scopo di indagare se esista un sistema di funzionamento familiare tipico, corrispondente allo sviluppo di tale sintomo in un membro del gruppo).

Questa situazione di lavoro ci costringe ad affrontare una serie di difficoltà particolari. Infatti nel tentativo di suscitare un vantaggioso mutamento in una famiglia si debbono cercare risposte a interrogativi sempre nuovi: che fare col professionista inviante che non sa nulla di terapia familiare, e si schiera coi genitori per tener fuori i bambini? Che fare se un membro della famiglia è in terapia individuale con qualcun altro? In che modo altre forme di terapia potrebbero essere associate alla terapia familiare? Quali dovrebbero essere le priorità dell'intervento? E quale è la « durata » migliore per la terapia familiare? A qual punto è opportuno terminarla? Problemi tutti che ancora attendono dai vari ricercatori risposte sufficientemente comprovate e convincenti.

La nostra équipe, composta, come piú sopra abbiamo scritto, di 5 terapeuti, 3 uomini e 2 donne, ha prescelto come « regola » di impiegare nel lavoro psicoterapico una coppia di coterapeuti di sesso diverso. La coterapia in coppia eterosessuale rappresenta un elemento fondamentale del nostro metodo di lavoro e di ricerca. L'introdurre nel gruppo familiare (la cui sofferenza è largamente collegata alla disfunzione del rapporto della coppia parentale) una coppia terapeutica relativamente matura, e collaborante in modo adulto, equivale a nostro avviso alle introduzioni di un modello piú elevato e completo di rapporto interumano, la cui complementarità reciproca, inerente alla funzionalità del rapporto, esclude idealmente le premesse della violenza nei suoi rispettivi aspetti di competizione, prevaricazione e resa passiva.

La coppia dei coterapeuti, inoltre, lavorando precipuamente sull'hic et nunc, cioè su quanto è direttamente osservabile nelle comunicazioni verbali e non verbali della seduta, e sul rapporto che la famiglia e i singoli membri volta a volta manifestano coi componenti la coppia, può giungere abbastanza rapidamente a rendere evidenti molte posizioni e difese profonde. Un tale evidenziamento è premessa per la rottura del sistema « omeostatico ». È infatti noto come il sistema familiare è tanto piú stabile quanto piú è patologico, « inchiodato » in un gioco senza fine che tende a ripetersi all'infinito. Tale omeostasi rigida è a nostro avviso piú difficilmente chiarificabile e modificabile da un terapeuta solo. Questi, infatti, diviene inevitabilmente unico oggetto di una serie molto complicata di proiezioni intricate e difficili da discernere. Mentre la coppia terapeutica permette ai vari membri della famiglia proiezioni piú differenziate, rispettivamente sul coterapeuta uomo e sulla coterapeuta donna, e quindi un gioco interattivo relativamente meglio specificabile. Terapeuti uomini e terapeute donne si scambiano inoltre in diverse combinazioni di coppia, allo scopo di permettere il confronto delle coppie fra loro, lo studio di eventuali variabili inerenti sia alla personalità e al rapporto dei coterapeuti che allo stile di lavoro di ciascuna coppia. Ciò tenendo sempre presenti le ben note difficoltà inerenti ad ogni coterapia.

Pensiamo interessante riferire una breve sintesi dell'attività sinora svolta, specie per i suoi riferimenti epidemiologici e socio-culturali. Giovani coppie in difficoltà di rapporto, senza prole, o con prole infantile asintomatica, prendono abbastanza facilmente contatto con noi in previsione di un trattamento. Il contrario avviene con le famiglie che presentano «pazienti designati », generalmente adolescenti o tardi adolescenti. È infatti assai piú difficile motivare queste famiglie alla psicoterapia, giacché la presenza di un membro psichicamente disturbato induce nei genitori, segnatamente nei padri, tali angosce, tosto fatte proprie dal « paziente designato », da rendere inaccettabile la proposta di una psicoterapia collettiva. I genitori preferiscono affidare il figlio « paziente designato » alla psicoterapia individuale piuttosto che affrontare la prevedibile messa in crisi delle proprie difese e l'evidenziamento dei conflitti irrisolti della coppia coniugale. Il vero e proprio accanimento con cui i membri delle famiglie disfunzionanti, nonostante la indubbia sofferenza, si adoperano per ostacolare qualsiasi mutamento dello status quo, mostra da un lato come il permanervi abbia in qualche modo una funzione rassicurante per l'intero gruppo, dall'altro come anche le strutture di personalità dei membri cosiddetti sani siano assai fragili. Infatti, per la prima volta nella storia del trattamento psichiatrico, invitando alle sedute esplorative, o alla psicoterapia congiunta, anche i membri della famiglia esenti da sintomi, si possono studiare direttamente quelli che mandano gli altri in psicoterapia, ma che mai la chiederebbero per sé, anche se, com'è di regola, sono parte integrante e pesante del sistema familiare patologico. Tali membri, nella nostra esperienza di lavoro, si rivelano gradualmente come i registi della resistenza al cambiamento. Da bravi registi, apparentemente collaboranti e volonterosi, non operano scopertamente sulla scena del dramma terapeutico, ma copertamente inducono gli altri ad agire i propri timori, le proprie resistenze, i propri impulsi alla fuga. Soltanto terapeuti sperimentati riescono a non cadere nel gioco, a smascherarlo con garbo, persuadendo il regista a uscire dalle quinte, a scoprirsi, a comunicare affettivamente con loro.

Qualche cifra aiuterà a presentare l'attività fin qui svolta nel nostro Centro (dati aggiornati al 31 marzo 1971).

Dal 1967 abbiamo esaminato 36 famiglie. Di queste:

3 coppie senza figli; 15 coppie con i figli in età infantile, apparentemente sani (i figli sono intervenuti alle sedute solo sporadicamente, o affatto); 18 famiglie al completo, per un totale di 540 sedute, la cui durata standard è di un'ora e un quarto: soltanto in circostanze eccezionali le sedute hanno avuto luogo al domicilio della famiglia in terapia; 18 famiglie su 36 hanno accettato la psicoterapia: 2 l'hanno interrotta dopo poche sedute, per le altre il trattamento è ancora in corso o è terminato con esito soddisfacente.

Delle 18 famiglie con cui non abbiamo oltrepassato la fase investigativa, 7 hanno interrotto le sedute prima di arrivare a una decisione terapeutica, 10 hanno rifiutato il trattamento proposto. A una famiglia il trattamento è stato sconsigliato. Cinque famiglie attualmente in psicoterapia hanno una paziente designata anoressica mentale.

Quanto alla nostra tecnica nel lavoro psicoterapico, dopo molti tentativi ed errori, ci siamo risolti ad abbandonare sempre piú il metodo interpretativo mirante a scoprire in un membro della famiglia, o nella famiglia come gruppo « le cause » del comportamento deviante; piuttosto siamo tesi a cogliere, e a mostrare « lo stile » della relazione. L'osservazione di certi comportamenti, soprattutto di quelli che presentano un carattere ripetitivo (ridondanze), ci permette un intervento diretto sui fenomeni, allo scopo di rilevarne effetti a significati in senso interazionale, mirando a impedire il loro ripetersi.

Come nella psicoanalisi, i commenti dei terapeuti possono anche in questo caso render conscio ciò che prima era stato inconscio, ma nella terapia familiare, a differenza che nella terapia psicoanalitica, si mira piuttosto come ha notato Wynne « a ciò che non è stato notato » ma è osservabile, che non « a ciò che non è stato notato », ma è deducibile.

Per esempio, nel corso di una tera-

Nathan W. Ackerman, recentemente scomparso, è il pioniere della psicoterapia della famiglia. È autore dei volumi *Psicodinamica della vita familiare* e *Patologia e terapia della vita familiare* pubblicati anche in Italia. Fondatore e direttore del ben noto Family Institute di New York, fu inoltre iniziatore della famosa rivista « Family Process ».

pia familiare in cui la « paziente designata » era una figlia adolescente, si poté osservare nell'arco di varie sedute un tipico comportamento ripetitivo. Ogniqualvolta la figlia, con visibile soddisfazione della madre, o anche abilmente stimolata da lei, acerbamente criticava il padre, la madre, che sedeva accanto al marito, allungava una mano a coprire dolcemente la sua.

L'aver rilevato l'aspetto ripetitivo di tale comportamento, e il suo effetto sui partecipanti alla transazione, non solo rese impossibile il ripetersi del fenomeno, ma permise di dimostrare con l'evidenza dei fatti l'intrico di alleanze segrete, continuamente formate, disciolte e negate, che costituivano un modo fondamentale nella disfunzione di quella famiglia, perpetuandovi la tensione dei conflitti e la confusione dei significati.

La nostra tecnica mira ad affinarsi per afferrare rapidamente le perturbazioni più evidenti del sistema di funzionamento transpersonale, con lo scopo di romperlo o capovolgerlo. È meno interessata a rendere cosciente la famiglia delle ragioni profonde (in senso psicodinamico) che l'hanno spinta e la spingono a scegliere quei modi di comportamento. In casi di interventi di urgenza in famiglie in crisi acuta, con pericolo imminente di ospedalizzazione del « paziente designato », tentiamo di annullare tale pericolo con interventi assai attivi che rompano il sistema ripetitivo, senza bisogno di evidenziare le singole dinamiche profonde che lo sottendono. Per esempio, in una famiglia di un adolescente gravemente caratteropatico, che presentava crisi frequentissime di violenza fisica che lo avrebbero in breve portato al ricovero, già in seconda seduta i terapeuti intuirono un'alleanza intima e segreta, mascherata dietro dispute accanite, tra il padre e il figlio, a spese della madre, angosciata ed esclusa. Nelle sedute successive il terapeuta uomo entrò rapidamente in campo, proponendo al figlio un modello virile provocatorio con cui venire, al limite alle mani, assai più affascinante di quello paterno, mentre la terapeuta donna inviava messaggi di comprensione alla madre esclusa. Con l'introduzione del terzo uomo, la diade padre-figlio fu scissa, arrivando così rapidamente alla rottura dell'omeostasi e alla cessazione del sintomo socialmente pericoloso.

In un'altra famiglia, la paziente designata era una bambina di nove anni, in condizioni estreme di emaciazione a causa del rifiuto di nutrirsi. Un precedente ricovero in una clinica neuropsichiatrica infantile, oltre a non aver sortito alcun effetto sulla gravità dello stato fisico, aveva seriamente traumatizzata la bambina dal punto di vista psichico, aumentando la sua determinazione a « punire » la famiglia, colpevolizzata « per averla messa in quel luogo orribile ». Allorché iniziammo la psicoterapia, il sistema funzionava in modo pseudoreciproco attraverso l'apparente schiavizzazione del resto della famiglia alla « paziente designata »: tutti erano costretti a mangiare enormemente affinché la bambina si nutrisse almeno un poco. Bastava però che la sorellina più piccola, esausta, lasciasse nel piatto un sol briciolo delle enormi porzioni che la paziente le imponeva, perché quest'ultima si astenesse totalmente dal cibo.

Alla sesta seduta, vista la gravità della situazione per lo stato di emaciazione estrema della paziente designata, nonché l'iperalimentazione pericolosa della sorellina, i terapeuti decisero un intervento di urgenza: d'autorità prescrissero alla piccola paziente un regime ridottissimo, ma sufficiente a vivere, e contemporaneamente ordinarono agli altri membri della famiglia di ingerire esattamente la stessa quantità di cibi e bevande ingerite dalla bambina. Ciò portò d'acchito al capovolgimento dell'aspetto più appariscente del sistema, alla mitigazione del sintomo più urgente e pericoloso come il rifiuto del cibo, e all'evidenziamento di nuovi fenomeni interattivi assai più importanti e significativi per il progresso terapeutico dell'intero gruppo familiare.

L'intervento terapeutico sulle famiglie in crisi acuta costituisce in questo momento il filone più importante della nostra ricerca e della nostra sperimentazione clinica. Se essa, nel momento attuale, si svolge in disparte delle pubbliche istituzioni psichiatriche, ha tuttavia la speranza di ritornarvi indirettamente con qualche frutto. Infatti, riuscire a costruire modelli teoretici e metodologici precisi, e come tali *trasmissibili*, per un rapido ed efficace intervento psichiatrico quale quello operato nei casi sopraccennati, potrebbe sensibilmente ridurre il numero dei giovani pazienti costretti a subire il trauma, talvolta irreparabile, di un ricovero istituzionale.

Nelle istituzioni poi, équipes allo scopo sperimentate potrebbero in futuro intervenire sulle famiglie e sull'ambiente dei giovani pazienti al primo ricovero, evitando così che il ritorno in un sistema « immodificato » abbia a perpetuare, come troppo spesso avviene, quell'alternarsi di ricoveri e dimissioni in cui tragicamente si configurano esistenze senza speranza.

La sorgente sacra della Senna

La fonte donde nasce il fiume era un tempo il santuario di Sequana, dea celtica della salute. Sul luogo è stata rinvenuta una eccezionale raccolta di sculture in legno che risalgono a duemila anni fa

di Simone-Antoinette Deyts

Molte civiltà antiche consideravano le sorgenti come luoghi sacri, forse perché il fenomeno dell'acqua che sgorga dal suolo sembrava loro magico. I greci e i romani praticavano i loro culti presso le fonti e prescrivevano agli ammalati che invocavano l'aiuto divino di bagnarsi in esse. Agli albori dell'era cristiana anche in Francia erano in uso simili riti. In particolare una fonte, situata vicino all'odierna città di Digione, era ritenuta sacra alla dea celtica Sequana e centinaia di pellegrini vi si recavano con la speranza di ritrovarvi la salute. Quella stessa fonte è la sorgente della Senna che attraversa Parigi e scorre fino al mare.

Un secolo di ricerche archeologiche alla sorgente della Senna ha portato alla luce i resti della struttura di un tempio, una quantità di offerte votive e una graziosa scultura in bronzo della dea Sequana. Durante gli ultimi dieci anni gli scavi compiuti sul terreno paludoso vicino al tempio hanno portato alla luce la piú grande raccolta omogenea di sculture celtiche in legno mai trovata: non meno di 400 effigi ben conservate, fra grandi e piccole. La maggior parte di esse sono immagini votive dedicate a Sequana. Fra le offerte vi sono effigi di pellegrini, piú grandi che al naturale, rozzamente intagliate da tronchi di quercia. La scoperta di queste sculture di legno porta un notevole contributo a quanto si conosce sulle arti dell'antica Gallia e alla comprensione delle pratiche religiose galliche: può essere perciò considerato l'avvenimento piú importante degli ultimi anni nel campo dell'archeologia celtica.

Un visitatore che lascia Parigi alla ricerca della sorgente sacra della Senna farà bene a evitare la strada principale per Digione e a prendere invece la strada per Troyes. Questo itinerario gli permetterà di fermarsi subito oltre Troyes, a Châtillon-sur-Seine, dove potrà cosí disfarsi di alcuni pregiudizi assai diffusi sull'antica Gallia. Qui egli potrà vedere il tesoro di Vix che comprende un diadema d'oro finemente cesellato e un vaso di bronzo alto all'incirca un metro e mezzo, ornato di un bellissimo fregio raffigurante dei guerrieri.

Questi due oggetti, insieme ad altri, vennero scoperti nel 1952 nella tomba di una principessa gallica. Di lavorazione greca, anziché locale, sono attribuibili al sesto secolo a.C. La loro presenza nella tomba reale sta a testimoniare una realtà della Gallia pressoché sconosciuta prima della conquista romana. In primo luogo, l'importazione dalla Grecia, dimostra che i galli avevano dei contatti commerciali con il mondo mediterraneo molti secoli prima che venissero i romani; e in secondo luogo, l'alto livello di queste opere prova la raffinatezza del gusto gallico in campo artistico.

Evidenze come queste sfatano la leggenda comune che i galli pre-romani fossero dei selvaggi avvolti in pelli e armati di terrificanti mazze. Questa immagine da aula scolastica del guerriero gallico ci viene tramandata dai romani, che riuscirono a vincere la resistenza gallica con gran difficoltà solo alla fine del primo secolo a.C. La falsità di questo luogo comune romano diverrà ancor piú ovvia se il visitatore percorrerà pochi altri chilometri verso sud sulla strada di Digione fino ad arrivare al bivio che porta alla sorgente della Senna. Qui troverà che il cosiddetto barbaro retaggio dei galli è caratterizzato non solo da un gusto squisito per l'arte, ma anche da evidenti testimonianze di una vita spirituale.

Non aspettatevi uno spettacolo grandioso. Ciò che vi apparirà sarà una valle che circonda un piccolo parco, ben tenuto e circondato da alberi e cipressi. Sulla destra si trovano una pozza e una grotta artificiali, accanto alle quali si erge un monumento innalzato nel 1867 dai parigini in omaggio al fiume che tanto ha contribuito alla vita della loro città. Questa pozza non è, tuttavia, la fonte sacra dei galli. Per trovare quest'ultima bisogna infatti attraversare il parco fino in fondo finché si incontra una bassa rupe che delimita da un lato la valle. Pochi metri piú avanti delle pietre segnano il posto. L'acqua sgorga fra i sassi, corre attraverso un canale rivestito di ciottoli fino a riempire un antico bacino artificiale dal quale straripa per confluire nel ruscelletto che proviene dal bacino piú recente.

Scrive il filosofo francese Gaston Bachelard: « L'acqua limpida rappresenta un innocente simbolo di purezza che attrae. L'uomo non ha bisogno né di cultura né di tradizioni per capire questa immagine naturale ». Questo simbolo sembra mantenersi vivo in tutte le epoche e in ogni luogo: infatti è un passo breve quello che porta dall'immagine della purezza al concetto di divinità che la presiede. Fortunatamente grazie al suo stato di conservazione possiamo capire in che modo i galli, che qui la veneravano, immaginassero Sequana, la personificazione divina della Senna. Poco prima del 300 d.C., probabilmente a causa di qualche invasore che imperversava nelle campagne, i custodi del tempio di Sequana rimossero l'immagine in bronzo della dea insieme a un'altra scultura e, per salvarle, le seppellirono in una stretta buca alle pendici della rupe. Il nascondiglio venne scoperto durante una ricerca archeologica nel 1933; al materiale ritrovato venne subito dato il nome di tesoro della Senna.

I galli rappresentarono Sequana in piedi su una barca, con le braccia aperte. I lineamenti della dea sono dolci e allo stesso tempo pieni di maestà; la sua posizione a braccia aperte sembra dare il benvenuto e offrire aiuto a chi le si avvicina. La barca che la porta è di una fattezza squisita: la prora è la testa di un'anatra che nel becco porta un frutto, la poppa è la coda eretta dell'anatra (*si veda la figura a pagina 49*); come si vede, è una barca del tutto appropriata per trasportare una dea del fiume. La scultura sembra essere stata eseguita durante il secondo secolo.

Grazie allo studio di generazioni di archeologi a partire dalla metà del diciannovesimo secolo, è ora possibile ricostruire ciò che doveva essere il santuario di Sequana e seguire lo sviluppo delle strutture del tempio dalle sue origini. All'incirca intorno al 50 d.C. venne costruito il canale di pietra che parte dalla fonte e venne eretto un piccolo tempio una decina di metri piú a nord, la cui architettura era tipicamente gallica e non aveva alcuna influenza romana. Era infatti costituito da una stanza quadrata, circondata da un portico aperto dove potevano raccogliersi i fedeli, che ospitava, con ogni probabilità, una immagine della dea. Evidenze archeologiche fanno supporre che alcuni dei riti di guarigione venissero officiati a un livello inferiore a ovest del tempio in una zona paludosa vicino alle rive dell'attuale ruscello. È possibile, inoltre, che la palude fosse un luogo rituale molti anni prima che venisse costruito il tempio.

Venne poi aggiunta una serie di costruzioni, probabilmente sotto influenza romana, a un livello superiore fra il bacino artificiale e le pendici della rupe. Le costruzioni piú a sud consistevano in sette stanze: una fila di quattro camere quadrate, all'incirca di quattro metri di lato ciascuna, e tre camere rettangolari piú grandi. Le quattro stanze piú piccole possono aver ospitato i sacerdoti che attendevano al santuario e le piú grandi possono essere servite loro per incontrarvi i pellegrini (per riunioni o forse per cure mediche).

Piú a nord un'imponente via d'accesso segnata da quattro pilastri conduceva i pellegrini in un lungo cortile. La parte est del cortile era occupata da un colonnato coperto che proteggeva un atrio il cui pavimento era di semplici mattoni, mentre il pavimento del cortile era tutto un mosaico a disegno geometrico. Sul lato nord del cortile tre scalini conducevano alla fonte, al canale e al bacino.

Questo grandioso complesso, costruito su di un'area lunga poco meno di un centinaio di metri, venne alla luce gradatamente durante un periodo di molte decine di anni. All'inizio non si poteva scorgere quasi nulla: quasi tutta l'opera in muratura, anche i singoli sostegni in pietra che avevano fatto parte dei pilastri del colonnato, erano stati portati via, durante i secoli successivi, per essere usati nelle costruzioni locali. Solo nel 1932 il lavoro condotto nella zona paludosa a ovest del tempio rivelò tre muri che delimitavano una zona rettangolare chiamata dagli archeologi « pozza ». Piú tardi altre ricerche rivelarono che quella zona, sebbene oggi paludosa cosí come lo era ai primi tempi del santuario, era stata riempita di terra poco dopo il 100 d.C. per farne un secondo cortile. Scavando in profondità la « pozza » ripiena di terra, nel 1963 e nel periodo 1966-67 si riportò alla luce l'eccezionale raccolta d'immagini votive di legno.

Queste importanti sculture non sono le sole offerte votive che siano state ritrovate alla sorgente sacra della Senna. In una delle stanze dei sacerdoti, per esempio, durante i primi scavi era stato rinvenuto un grande boccale in terracotta che portava la seguente iscrizione in latino: « *Deae Sequana(e) Rufus donavit* » (« Rufo diede questo alla dea Sequana »). All'interno del boccale ve n'era un altro piú piccolo che conteneva piú di 800 monete di bronzo del terzo e quarto secolo. Intorno al boccale piú piccolo vi erano poi 120 minuscole placche votive di bronzo. È probabile che il Rufo dell'iscrizione non avesse niente a che fare con ciò che il boccale conteneva al momento della sua scoperta; infatti 120 offerte sono difficilmente attribuibili a una sola persona e, inoltre, molte delle placche descrivevano malattie femminili. Sicuramente i due boccali e il loro contenuto devono aver fatto parte del tesoro di un tempio che era stato rubato e seppellito temporaneamente al sicuro dai predoni durante un periodo di disordini.

L'iscrizione convenzionale latina sulle offerte votive è VSLM, un'abbreviazione di « *Votum solvit libens merito* » (« Offre con piacere il dovuto omaggio », o, piú idiomaticamente, « Il tal dei tali adempie al suo voto »). Questa iscrizione manca dal boccale di Rufo, ma appare su molti altri oggetti ritrovati al santuario di Sequana. « *De Sequana* », si legge su una delle offerte « *Sienulla Vectii filia VSLM* » (« Sienulla, figlia di Vezio, adempie il suo voto a Sequana »).

Evidentemente i galli pensavano di

Queste due sculture in legno di chiara influenza romana, ritraggono un giovane (*a sinistra*) e una giovane (*a destra*) che vissero in Gallia all'incirca al tempo di Cristo. Sono state ritrovate recentemente al santuario di Sequana insieme ad altre 400 sculture in legno. L'ottimo stato di conservazione dipende dal fatto che sono rimaste sotterrate in terreno umido.

Il santuario di Sequana, costruito accanto alla fonte della Senna (*a*), confina con un monumento moderno (*in basso*) inalzato dagli abitanti di Parigi nel 1867. In origine, il santuario era composto da un bacino artificiale (*b*) sotto la fonte (*a*) e da un semplice tempio nelle vicinanze (*c*). Piú tardi, probabilmente sotto l'influenza romana, venne aggiunto un cortile (*d*), con delle stanze per i sacerdoti a un'estremità (*e*) e, su di un lato, un portico coperto. Il « tesoro della Senna », che comprende una scultura di Sequana, venne nascosto nei pressi della rupe subito a est del sito (*f*). Oggetti in legno vennero gettati nella zona paludosa (*g*), in seguito bonificata.

onorare maggiormente la dea rivolgendosi a lei nella lingua dei loro conquistatori. Sono state ritrovate, alla sorgente della Senna, undici iscrizioni latine, sebbene in alcuni punti il latino sia un poco approssimativo. In altre iscrizioni sono stati usati i caratteri latini, ma solo per tradurre il linguaggio locale. Una iscrizione di cinque righe sulla faccia di un'offerta votiva appare scritta in caratteri latini, mentre sul retro è in caratteri greci; su ambedue i lati le parole sono puramente di origine gallica.

Finora 1000 offerte votive in pietra o in bronzo sono state ritrovate al santuario di Sequana. Quelle in pietra, ricavate da una tenera pietra calcarea che si trova sul posto, sono le piú numerose. Generalmente rappresentano la parte ammalata del corpo di chi faceva l'offerta, cosicché vi si ritrovano le immagini di braccia, gambe, mani, piedi, teste, seni e organi sessuali, e anche alcuni tentativi di rappresentare l'anatomia interna. Generalmente nessuno sforzo veniva fatto per indicare la natura della malattia, sebbene alcuni medici affermano che certe immagini ritraggono la gotta, l'ernia o anche il ritardo mentale.

In aggiunta a queste rappresentazioni di una sola parte del corpo vi sono anche busti e figure intere in pietra, alcune di queste alte fino a un metro. L'individuo cosí rappresentato, che implorava la salute o semplicemente venerava Sequana, è raffigurato quasi sempre nell'atto di recare un'offerta: un cane, un coniglio, della frutta o una borsa piena di denaro (*si veda la figura a pagina 48*).

Le placche in bronzo sono, invece, una cosa del tutto diversa. Sono dei sottili pezzi di metallo, quasi tutti larghi meno di 5 cm. Pare evidente che venissero prodotti in serie, ritagliandoli da una lamiera, man mano che ve ne era bisogno; la raffigurazione di una parte del corpo veniva poi incisa sulla piastra. Se occorreva raffigurare gli occhi di un pellegrino, venivano fatti dei contorni ovali o rotondi; in alcuni casi vi si aggiungevano le pupille e le ciglia. Gli occhi, in alcune piastre, sono dei semplici buchi, forse a sottolineare che il donatore era cieco (*si veda la figura in alto a sinistra a pag. 49*).

Gli organi sessuali e il seno sono le altre parti anatomiche che appaiono piú frequentemente sulle piastre. Anch'essi venivano stilizzati. Il bacino generalmente rappresentato da un triangolo: quello maschile veniva distinto da un fallo, mentre quello femminile veniva lasciato senza segni. Le raffigurazioni del seno erano le stesse, anche se i donatori, certe volte, aggiungevano dei dettagli.

Quando gli scavi della « pozza » rivelarono gli oggetti di legno preservati dal suolo acquitrinoso, fummo fortunati nel poter usufruire dell'esperienza di archeologi che avevano studiato vari modi per conservare simili reperti. Se il legno antico impregnato d'acqua non viene protetto, asciugandosi si riduce in polvere. Le due tecniche principali per il suo trattamento sono l'essiccazione mediante congelamento e la sostituzione dell'acqua; scegliemmo questo secondo metodo perché piú conveniente trattandosi di un gran numero di pezzi di dimensioni molto diverse. Come si rinvenivano gli oggetti, essi venivano immersi in una soluzione acquosa di una cera solubile (glicol polietilenico). Quando la soluzione ha il giusto grado di concentrazione, gradualmente si sostituisce all'acqua che si trova in ogni cellula del legno della statua. Dopodiché si può esporre il legno all'aria con tutta tranquillità. Ricevemmo inestimabili aiuti e consigli per l'applicazione di questa tecnica di conservazione dall'Istituto reale del patrimonio artistico di Bruxelles. Alcune delle piú grandi sculture in legno, come pure quelle in pietra della stessa grandezza erano ritratti di pellegrini, tuttavia si riscontra tra loro una gran differenza: le sculture in pietra sono dolci e armoniose nei contorni; quelle in legno, sebbene in nessun modo artisticamente inferiori, sono invece rigide e spigolose. Molte di queste ultime rappresentano figure umane con mantello e cappuccio. Le pieghe degli indumenti nascondono le braccia tenute strette al corpo; guardandole se ne ricava un'impressione di forza e di quieta tenacia. In particolare, la figura di un uomo incappucciato commuove per la sua goffaggine e per i suoi piedi rivolti all'interno, fedelmente riprodotti.

Il mantello con cappuccio è un indumento locale, il *bardocucullus* che il poeta latino Marziale prendeva in giro. Quando il cappuccio non ricopriva la testa, veniva gettato indietro e il mantello cadeva dritto con poche pieghe. Si pensa che questo fosse un indumento da cerimonia, introdotto all'inizio dai bardi della Gallia, e che in seguito acquistò un significato religioso.

Vi sono altri tipi di vestiario che ornano le sculture in legno. Uno di questi è un semplice abito diritto, senza alcun dettaglio al collo, che scende alle caviglie ed è tenuto in vita da una cintura larga e pesante. Questo tipo di abito appare frequentemente. Si trovano pure tuniche di varie fogge e lunghezze; una, in particolare, ricorda in modo incredibile una corta giacca moderna.

Le statue di donne ritrovate non sono sufficienti per poter avere un'idea generale sull'abbigliamento femminile, ma, giudicando da un esempio abbastanza dettagliato, un abito femminile comune potrebbe essere stato una tunica lunga fino ai piedi e con una scollatura quadrata. La tunica era portata sotto un mantello, aperto sul seno, che cadeva dalle spalle al ginocchio sul davanti e continuava con una linea obliqua sul dietro.

La più bella delle sculture femminili rivela poco o niente sull'abbigliamento. Il materiale usato dallo scultore era un ramo d'albero; egli si è limitato a modellare solo le due estremità, aggiungendo pochi particolari per tutta la lunghezza del legno. Una piccola curva suggerisce le spalle; un'altra curva, quasi impercettibile, situata in fon-

L'uomo dal mantello con cappuccio è una delle tante sculture in legno che ritraggono delle figure intere. La posizione dei piedi rivoltati all'interno forse indica il difetto fisico di cui soffriva il donatore.

Offerte votive di tre tipi diversi sono, da sinistra verso destra, un braccio, l'interno di un corpo umano che mette in evidenza lo stomaco, e una serie di teste. Chi fabbricava queste teste votive le vendeva, probabilmente, una per volta ai pellegrini.

do alla figura fa intuire il bordo inferiore della veste. Sotto di essa, incredibilmente piccoli, appaiono i piedi.

Il viso della donna, però, colloca questo lavoro fra i piú belli fra quelli in legno. I capelli, divisi nel mezzo, cadono lisci giú per il collo. La fronte è alta, la linea delle sopracciglia marcata, il naso ben delineato, il labbro superiore corto e quello inferiore imbronciato. Il viso ovale e tutte le sue parti si armonizzano cosí da presentare una immagine piena di forza che è, allo stesso tempo, intensa e controllata (*si veda la figura in questa pagina*).

Tra le offerte votive in legno si ritrovano delle strane sculture consistenti in due, tre o perfino quattro teste, scolpite l'una sopra l'altra su un solo ceppo. Queste ci ricordano all'istante i totem indiani dell'America nord-occidentale. La loro fattura è rozza e geometrica. Una linea diritta indica la fronte e gli occhi, la bocca è rappresentata da una piccola fessura e il naso da una piramide.

Che cosa volevano significare queste immagini ripetute? Esaminandole si scopre che vi sono due tipi di ripetizioni. In alcune di queste sculture il legno che separa una testa dall'altra è inciso profondamente, suggerendo, cosí, che lo scultore intendeva separare le teste. In altre, invece, le teste non hanno praticamente collo cosicché non vi è spazio per separarle. Le prime, quindi, ricordano le placche in bronzo prodotte in serie: erano probabilmente delle offerte individuali preparate in gran numero con un minimo di lavoro e di spreco; su richiesta, ogni testa poteva venir separata. Le seconde sembrano aver avuto uno scopo diverso; forse hanno un significato analogo alle preghiere ripetute piú volte. Se Sequana veniva invocata nel modo appropriato, ella avrebbe rivolto la sua sollecita attenzione al pellegrino la cui testa veniva rappresentata da un'offerta votiva. Se la testa veniva riprodotta piú volte, i suoi aiuti non si sarebbero forse moltiplicati in proporzione?

Questa pratica, comune alle religioni primitive ancora esistenti, ci viene suggerita anche da altre tracce ritrovate nello stesso luogo. Fra le offerte votive in pietra ve ne sono alcune che ritraggono tre teste o sei gambe. Stranamente, in queste sculture le ripetizioni sono orizzontali piuttosto che verticali. Probabilmente la sequenza orizzontale era piú economica usando la pietra, cosí come quella verticale lo era per il legno. I problemi materiali, non meno di quelli spirituali, obbediscono a certe leggi e i galli erano, certamente, un popolo pratico.

Due teste in legno, che non sono state certo prodotte in serie, ci dicono qualcosa sull'influenza artistica che ebbero i romani sui galli in quel periodo. Una di queste raffigura un giovane e l'altra una fanciulla (*si veda la figura a pagina 42*). La testa maschile è grande; la sua circonferenza dipende evidentemente dalle dimensioni del ceppo da cui è tratta. Una curva continua corre dalla cima del capo ai capelli sulla nuca; sulla fronte i capelli formano una frangia. L'arco sopraccigliare sporgente ripara gli occhi con folte ciglia ben delineate. Il naso è grande; due linee profonde ai lati delle narici scendono fino agli angoli di una piccola bocca carnosa sopra un mento arrotondato. I dettagli di questo lavoro sono fantastici. La viva cesellatura della bocca e del mento contrastano con la levigatezza delle altre parti del viso cosí da accentuarne la forza.

Un grosso ramo d'albero è stato trasformato in una figura di donna scolpendone le due estremità e lasciando il resto intatto. A sinistra si vede la figura per intero; a destra, la sola testa, quasi di profilo.

Il viso della giovane donna è lungo e magro. La fronte è bassa, gli occhi hanno la palpebra pesante, il naso è diritto e forte, la bocca sottile e con le labbra serrate. I capelli, in grosse trecce, scendono dalla sommità del capo fino al collo, incorniciando il viso; le trecce sono modellate in bassorilievo. Lo scultore era condizionato nel suo lavoro dalle strette dimensioni del tronco, ma non per questo il suo ritratto, con quel suo senso di equilibrio e di serenità, risulta meno interessante.

Questi due studi si allineano con le piú belle fra le sculture galliche, o piú propriamente gallico-romane. Quale è la loro collocazione storica? Le ceramiche e le monete riportate alla luce insieme alle offerte votive in legno fanno pensare che siano state sotterrate intorno al 50 o al 60 d.C.

Devo però sottolineare che questa data non è del tutto certa; il lavoro di scavo della palude non è ancora terminato e altre scoperte potrebbero cambiare questa datazione preliminare. Tuttavia, supponendo che sia corretto fissare la data del loro sotterramento verso la metà del primo secolo d.C., ne consegue che le sculture sono state eseguite antecedentemente, forse anche prima che i romani giungessero in questa zona della Gallia, nel tardo primo secolo a.C. La natura « locale », non raffinata, di molti fra gli oggetti ritrovati nella palude tende a convalidare questa ipotesi.

Allo stesso modo i due ritratti provano che l'influenza, se non la presenza romana, veniva sentita in Gallia sin da data remota. Ritratti romani di teste, molto rassomiglianti a questi, ma solo in bronzo, sono stati ritrovati sia in Francia che in Svizzera. La rassomiglianza sembrerebbe indicare che gli scultori del santuario di Sequana devono averne visti degli esemplari. Sembra che non ci sia voluto molto tempo agli scultori gallici per assimilare la lezione che veniva loro dall'estero. Essi applicarono la loro abilità nel modellare dei ritratti viventi non in bronzo, ma in legno, materiale di cui avevano completa padronanza.

Possiamo domandarci come apparisse il santuario di Sequana all'epoca in cui era un centro di pellegrinaggio gallico. A questo proposito faremmo bene a ricordarci del luogo di guarigione famoso in Grecia, visitato dal geografo Pausania nel secondo secolo della tarda era cristiana: il tempio di Esculapio a Epidauro, oltre Atene, al di là del golfo Saronico. La descrizione di Pausania del santuario greco suggerisce delle interessanti analogie con quello gallico. I pellegrini che chiedevano aiuto

Un ignoto artista da un blocco di legno di quercia di limitate dimensioni ha scolpito questa realistica e possente immagine vagamente deforme di un robusto toro.

Questa figura in pietra di pellegrino testimonia che questo materiale, generalmente calcareo, era molto usato dai fabbricanti di offerte votive. Gli oggetti scolpiti nella pietra superano per numero ogni altro tipo di offerte fatte a Sequana. I pellegrini sono ritratti nell'atto di portare doni; in questo caso si tratta di una borsa di denaro.

a Esculapio dovevano passare una notte nei recinti del tempio. Era d'obbligo una precedente purificazione; infatti la presenza di tubazioni d'acqua vicino al portico del tempio ci fa pensare che il rito richiedesse un'abluzione. In ogni modo i pellegrini passavano la notte sotto il portico del tempio; ad alcuni di essi appariva in sogno Esculapio, e i fortunati sognatori potevano considerarsi guariti. Era poi costume fare un'offerta per l'incisione di una iscrizione sul tempio, nella quale era riportato il nome del pellegrino, la descrizione della malattia che lo aveva spinto a venire e la natura della guarigione. Pausania vide sei enormi pilastri ricoperti di iscrizioni di questo tipo durante la sua visita a Epidauro.

In termini di paragone fra Epidauro e il santuario di Sequana, è chiaro che qui non mancava certamente l'acqua per le abluzioni religiose. Il bacino artificiale sotto la fonte aveva quasi sicuramente fatto parte del santuario anche prima che venissero costruiti il colonnato e il portico. Sembra probabile che, una volta costruita questa parte del santuario, la purificazione nel bacino venisse seguita da una veglia notturna nel portico. I sacerdoti, dalle stanze vicine, sarebbero stati pronti ad aiutare fornendo spiegazioni, se Sequana si fosse rivelata in sogno a qualche pellegrino.

È sostanzialmente provato dal numero delle offerte votive che i pellegrini gallici avevano un loro modo caratteristico di commemorare una visita fortunata a Sequana. Questa consuetudine sopravvisse in altre parti della Francia fino al sesto secolo, come è dimostrato dalla descrizione dovuta a un notabile della Chiesa, Gregorio di Tours, degli scempi compiuti in un tempio locale; Gregorio scrisse: « I barbari vicini vennero a compiere dei sacrifici e a riempirsi di carne e vino sino al punto di vomitare; lí adoravano idoli come divinità e depositavano sculture in legno che rappresentavano una parte del loro corpo, affetto da qualche male. » Gli antichi costumi sembra siano stati duri a morire. Anche i primi concili ecclesiastici tuonavano contro queste pratiche pagane. « Distruggete le fonti e i boschetti che osano chiamare sacri », ordina un decreto del sesto secolo. « Proibite le immagini che vengono appese agli incroci e, se ne trovate, bruciatele. Siate certi che non verrete guariti in altro modo se non pregando la croce di Cristo ».

Dove potevano essere acquistate le offerte votive? La presenza al tempio di Sequana di oggetti prodotti in serie, fa pensare che nelle vicinanze vi fosse una via con negozi che vendevano tali oggetti. L'esistenza di un simile commercio aiuterebbe a far capire lo strano fenomeno della diversa qualità artistica delle offerte votive: infatti su una dozzina di sculture di nessun conto se ne trova una veramente raffinata. Indubbiamente il diverso valore delle offerte votive dava modo ai negozianti di servire pellegrini dalle diverse possibilità economiche.

Con la fantasia possiamo facilmente far rivivere la vallata di Sequana. Vi sarà una moltitudine di gente; non solo gli ammalati, ma i loro amici e parenti. I malati, forse, mostreranno la loro devozione indossando il semplice mantello da cerimonia. L'abbigliamento dei loro compagni invece potrà essere tanto fantasioso quanto lo desideri l'immaginazione. Vi sarà ressa ai negozi per contrattare l'acquisto delle offerte votive. Nel cortile, pavimentato a mosaico, saranno disposti dei tavoli da questua che metteranno in mostra gli acquisti dei pellegrini. All'interno del piccolo tempio vi saranno delle panche gremite dove verranno esibite le testimonianze di guarigioni passate. Qui, quasi certamente, le offerte votive piú importanti occuperanno una posizione preminente, mentre le piú umili verranno relegate in qualche angolo oscuro.

Con il passar del giorno diminuirà il numero dei pellegrini. I fedeli, che non avranno ancora potuto parlare con i sacerdoti, si dirigeranno verso le vicine locande aspettando il mattino seguente. Cosí faranno anche i compagni dei pellegrini scelti quel giorno, mentre questi ultimi si laveranno prima di entrare sotto il portico di Sequana per passarvi la notte.

I ciechi vengono guidati, gli storpi assistiti, forse una parte del portico è riservata a gruppi di donne. Alla fine, scende la notte e arriva l'ora in cui Sequana manterrà la promessa delle sue braccia aperte. E se a questo punto lasceremo che la realtà prenda il posto dell'immaginazione, ci accorgeremo che il vero tesoro della Senna è costituito dalle semplici rovine del tempio dinanzi a noi e da ciò che queste ci fanno intravedere del passato abbastanza ignoto dell'antica Gallia.

Le placche di bronzo (*a sinistra*) sono il terzo tipo di offerte votive ritrovate al santuario di Sequana. Con ogni probabilità, le placche venivano offerte da pellegrini sofferenti agli occhi. Quella che raffigura delle orbite vuote (*la prima a sinistra*) sta forse ad indicare la cecità dell'offerente. Il bronzo di Sequana mostra la dea in piedi su una barca dalla prora fatta a becco di anatra e dalla poppa che termina con la coda dell'animale. La figura è interpretata nel piú tradizionale stile romano.

Una nuova classe di laser a diodo

"Eterostrutture" minute e a forma di sandwich composte di due o piú materiali semiconduttori differenti sono promettenti quali generatori di onde portanti a buon mercato per le comunicazioni di massa

di Morton B. Panish e Izuo Hayashi

Subito dopo l'invenzione del laser circa una decina d'anni fa, si riconobbe che una delle piú importanti applicazioni potenziali di questo nuovo strumento sarebbe stata quella di un generatore di onde portanti per un sistema di comunicazioni ottiche a lunga distanza. In linea di principio, un singolo fascio laser possiede una capacità di portare un segnale molte migliaia di volte piú grande di qualsiasi altro mezzo di trasmissione in servizio allora o adesso. Purtuttavia, l'applicazione commerciale su larga scala del laser nelle comunicazioni non si è ancora verificata, e potrebbe non verificarsi per un certo tempo. Questa situazione è naturalmente in parte il risultato del fatto che una tale applicazione richiede lo sviluppo di una vasta e complessa tecnologia associata. In aggiunta comunque c'è stato il fatto che la maggior parte dei laser a disposizione è stata finora troppo grande, troppo delicata, troppo costosa e troppo complicata per competere economicamente con le attuali tecniche di comunicazione di massa. L'unica eccezione a molti di questi inconvenienti è rappresentata da laser a diodo semiconduttore.

Il primo successo nell'ottenere luce laser da diodi semiconduttori fu annunciato quasi simultaneamente nel 1962 dai gruppi che lavoravano alla General Electric Company, alla International Business Machines Corporation e ai Laboratori Lincoln dell'Istituto di Tecnologia del Massachusetts. Questi tre gruppi raggiunsero il successo con diodi fatti del composto semiconduttore arseniuro di gallio. Il primo annuncio fu salutato con considerevole entusiasmo dai ricercatori del settore, poiché i nuovi laser, a differenza dei loro piú grandi predecessori, potevano essere molto piccoli, piuttosto robusti, a buon mercato e di semplice impiego. Ci si aspettava che essi lavorassero, almeno in linea di principio, a una efficienza piuttosto alta in confronto agli altri laser. Inoltre essi richiedevano solo una tensione di lavoro da uno a due volt. In realtà è possibile, almeno in certe applicazioni, paragonare i laser piú convenzionali e grandi al laser a diodo, pressappoco nella stessa maniera in cui si potrebbero comparare i tubi a vuoto coi transistor.

I primi laser a diodo, comunque, avevano un notevole svantaggio. Essi per operare richiedevano una densità di corrente talmente alta che, per proteggerli contro i danneggiamenti risultanti dal surriscaldamento, non potevano essere utilizzati con continuità a eccezione di quando lavoravano a temperature molto basse. Era possibile farli funzionare a impulsi cortissimi (approssimativamente un milionesimo di secondo) a temperatura ambiente, ma questo modo di operare limita notevolmente l'utilità di un tale dispositivo sia nelle comunicazioni sia nelle altre applicazioni.

Nella discussione che segue sono descritte le proprietà di una nuova classe di laser a diodo semiconduttore che può essere utilizzata con continuità a temperatura ambiente. La evoluzione di questa nuova classe di laser – chiamati laser a eterostruttura – porta piú vicino alla realtà il raggiungimento di un sistema competitivo di trasmissioni ottiche per telecomunicazioni.

Il laser a diodo è un perfezionamento del diodo a elettroluminescenza. Un tale dispositivo produce luce in una maniera piuttosto differente da quella delle ben note lampade a incandescenza. Nelle lampade a incandescenza il filamento di tungsteno riscaldato elettricamente emette luce su una singola banda di lunghezza d'onda ampia un migliaio di angstrom poiché una parte della energia di eccitazione termica degli atomi è trasferita ai fotoni che assumono un ampio intervallo di energie caratteristiche della temperatura del filamento emettitore. I diodi a luminescenza producono una stretta banda di radiazioni, generalmente di ampiezza inferiore ad alcune centinaia di angstrom con un picco nelle lunghezze d'onda caratteristiche del materiale luminescente. I diodi a luminescenza non richiedono aumenti nella temperatura; infatti la loro efficienza cresce quasi sempre quando la loro temperatura diminuisce. L'elettroluminescenza di questi dispositivi è un risultato diretto delle proprietà dei materiali semiconduttori.

I cristalli semiconduttori, come ogni altro cristallo, sono materiali in cui i singoli atomi sono legati l'uno all'altro in un ordine tridimensionale ripetuto. Un tale ordine si dice reticolo cristallino. I legami tra i singoli atomi del reticolo sono soprattutto basati sull'interazione degli elettroni piú esterni degli atomi che costituiscono il reticolo. Questi elettroni, oltre che a tener legato assieme il cristallo, conferiscono al materiale le sue caratteristiche proprietà ottiche. In un atomo isolato (per esempio un atomo di un gas) ogni elettrone può avere ciascuna di un certo numero di energie distinte. Queste energie sono chiamate « livelli » e un elettrone dotato di una particolare energia è descritto come occupante un particolare livello di energia. Il livello che un elettrone occupa dipende dalla sua posizione nella struttura elettronica dell'atomo, dalla temperatura, dal grado di eccitazione esterna e da altri fattori.

La struttura dei livelli energetici di un atomo isolato può infatti essere descritta come una specie di scala a pioli (*si veda la figura in alto a pag. 52*). Per un atomo in un basso stato di eccitazione i « pioli » (o livelli) piú alti sono tutti vuoti. Aggiungendo energia in quantità discrete gli elettroni possono essere fatti saltare da un piolo a uno piú alto, in modo che nei piú alti stati di eccitazione dell'atomo alcuni dei pioli piú bassi sono vuoti e alcuni dei piú alti sono occupati. La luminescenza è dovuta al trasferimento dell'energia di eccitazione dall'elettrone a un quanto discreto di luce, o fotone, quando l'elettrone passa da un livello piú alto a uno piú basso. Il fotone può essere pensato come un'onda con una data lunghezza d'onda oppure come una particella con una data energia.

Nei cristalli i livelli individuali di energia degli elettroni esterni degli atomi isolati si mescolano fino a formare bande quasi continue di livelli energetici disposti molto vicini. I livelli in queste bande possono essere considerati come un continuo di stati di energia, ciascuno dei quali può essere occupato da un elettrone. Il numero di stati in una banda è uguale al numero di atomi nel cristallo moltiplicato per un numero intero piccolo (in genere minore di 10). Si descrivono le bande come vuote, piene o parzialmente piene a seconda che nessuno degli stati sia occupato da elettroni, tutti siano occupati da elettroni o soltanto alcuni di essi siano occupati da elettroni. Le bande sono separate l'una dall'altra da regioni di energia « proibite ». Salvo speciali circostanze gli elettroni del cristallo non possono avere energie nell'intervallo proibito.

Per gli scopi di questa discussione è sufficiente considerare un tipico semiconduttore come se avesse due bande di energia di questo tipo. Quella a energia piú alta, chiamata banda di conduzione, è separata da quella a energia piú bassa, chiamata banda di valenza, da un intervallo di energia proibita avente un'ampiezza che indicheremo con E_g. La grandezza di E_g ha un effetto di controllo sulle proprietà ottiche del semiconduttore. In generale il semiconduttore sarà trasparente alla luce avente fotoni di energia inferiore a E_g e assorbirà la luce avente fotoni di energia maggiore. La luce con fotoni di energia prossima a E_g è assorbita in misura maggiore o minore in dipendenza di molti fattori ma particolarmente in dipendenza della quantità dei vari atomi estranei presenti quali impurezze nel cristallo.

La banda di conduzione non contiene quasi alcun elettrone a meno che il cristallo non contenga impurezze « donatrici ». La banda di valenza è pressoché riempita di elettroni. Essa ha un piccolo numero di stati vuoti, chiamati buche, che agiscono come fossero par-

Il nuovo laser progettato dagli autori nei Laboratori della Bell Telephone è mostrato qui nella cruna di un ago per sottolineare le sue dimensioni estremamente piccole. Il dispositivo, chiamato laser a diodo a eterostruttura doppia, è il primo laser a diodo semiconduttore capace di operare con continuità in condizioni di temperatura ambiente.

STRATO 4
STRATO 3
STRATO 2
STRATO 1
SUBSTRATO

La sezione trasversale del nuovo laser appare in questa micrografia a scansione elettronica di una faccia di clivaggio del dispositivo. La doppia eterostruttura mostrata in questa figura è essenzialmente composta da una grande quantità di strati semiconduttori molto sottili e « drogati » da impurezze, cresciuti su un substrato di arseniuro di gallio. Lo strato 1 è arseniuro di gallio e alluminio drogato con stagno. Lo strato 2, la regione attiva per il lasing, è arseniuro di gallio drogato con silicio e zinco. Lo strato 3 è arseniuro di gallio e alluminio drogato con zinco. Lo strato 4, che serve come strato di contatto, è costituito solo da arseniuro di gallio drogato con germanio.

ticelle cariche positivamente. Queste particelle immaginarie sono dovute in primo luogo alla presenza di atomi « accettori » estranei, che hanno accettato elettroni dalla banda di valenza, lasciando gli stati vuoti, o buche. Le buche nella banda di valenza e gli elettroni nella banda di conduzione sono le sole entità presenti nel semiconduttore che portano corrente. Se vi sono più donatori che accettori, la corrente nel materiale è portata da elettroni della banda di conduzione (carichi negativamente) e il materiale è designato come semiconduttore di tipo *n*. Se vi sono più accettori che donatori, la corrente è portata dalle cariche positive e il materiale è designato come semiconduttore di tipo *p* (*si veda l'illustrazione in basso in questa pagina*).

In questi due diagrammi sono confrontati gli stati di energia permessi di un elettrone in un atomo isolato (*a sinistra*) e in un solido semiconduttore (*a destra*). In un atomo, un elettrone (*punto nero*) può occupare uno fra parecchi livelli discreti di energia in dipendenza del suo grado di eccitazione. Quando l'elettrone assorbe energia, si muove verso un livello più alto (*freccia verso l'alto*); quando perde energia, cade a un livello più basso (*freccia verso il basso*) e nel processo emette un fotone di luce la cui lunghezza d'onda è inversamente proporzionale alla differenza di energia tra i livelli (*freccia colorata*). Questo fenomeno è chiamato luminescenza. In un solido cristallino, invece, i livelli energetici si fondono per formare bande, ciascuna delle quali contiene approssimativamente un numero di livelli pari al numero degli atomi nel cristallo. Queste bande sono separate l'una dall'altra da un intervallo di energia proibita (E_g) entro la quale gli elettroni del cristallo non possono esistere. Un semiconduttore puro ha una coppia di bande di energia adiacenti; quella più bassa, chiamata banda di valenza, è usualmente riempita da elettroni, mentre quella più alta, chiamata banda di conduzione, è usualmente vuota. In un cristallo semiconduttore la corrente elettrica è portata soltanto dalle cariche che stanno in bande parzialmente occupate, quali quelle che si ottengono quando un elettrone è eccitato dalla banda di valenza alla banda di conduzione attraverso l'assorbimento di energia pari all'intervallo tra le bande, sia termicamente che attraverso qualche altro mezzo di eccitazione. Questo processo lascia non occupato un livello di energia, o « buca » (*cerchietto*), nella banda di valenza. Sia l'elettrone eccitato carico negativamente nella banda di conduzione sia la buca carica positivamente nella banda di valenza possono agire come portatori di corrente. La conducibilità elettrica di un semiconduttore è normalmente piuttosto bassa, poiché solo un numero comparativamente piccolo di cariche sono create termicamente. La luminescenza ha luogo in un semiconduttore quando un elettrone eccitato nella banda di conduzione « si ricombina » con una buca mobile della banda di valenza.

Atomi di impurezze con livelli di energia nella regione di separazione tra le bande di un materiale semiconduttore possono essere usati per migliorare la conducibilità elettrica del semiconduttore. Atomi donatori (*a sinistra*) forniscono alla banda di conduzione un elettrone, creando cariche positive immobili ai livelli donatori. Atomi accettori (*a destra*) accettano un elettrone dalla banda di valenza, lasciando una buca.

La creazione di coppie di elettroni e buche in materiali di tipo *n* o *p* è possibile in diversi modi. Queste cariche extra sono presenti in condizioni di « non equilibrio » e il cristallo tende a eliminarle. L'eliminazione delle cariche iniettate provoca la perdita dell'energia di eccitazione da parte dell'elettrone eccitato. Poiché l'eccitazione dell'elettrone dalla banda di valenza alla banda di conduzione implica la cessione dell'energia E_g, questa è la quantità di energia che esso può perdere. L'elettrone « rioccupa » allora la buca iniettata nella banda di valenza in un processo chiamato di ricombinazione buca-elettrone. Fra le maniere con cui questa energia può essere persa vi è l'emissione di un fotone (luminescenza). Questo è il processo di cui ci siamo occupati inizialmente. Esso si manifesta con notevole efficienza nei semiconduttori quali l'arseniuro di gallio.

Quando parte di un cristallo semiconduttore è di tipo *n* e parte è di tipo *p*, il confine fra le due regioni viene detto giunzione *p-n*. Tale giunzione forma la base dell'effetto transistor e anche dell'azione dei diodi luminescenti. Poiché una giunzione *p-n* è una superficie di confine che ha cariche negative mobili (elettroni) da una parte e cariche positive mobili (buche) dall'altra, ci si può aspettare che, a causa dell'ordinario moto termico, queste cariche debbano muoversi verso la giunzione, dove esse dovrebbero ricombinarsi, generando nel diodo un flusso di corrente. In realtà questo processo ha inizio, ma si arresta dopo che un piccolo numero di buche e di elettroni hanno raggiunto la giunzione e si sono ricombinati, poiché le buche mobili e gli elettroni hanno lasciato indietro centri di impurezze costituiti da cariche positive e da cariche negative immobili. Il campo elettrostatico dovuto alle cariche immobili si oppone al moto e arresta il flusso di corrente. Come risultato si ottiene che le bande di valenza e di conduzione si deformano (*si veda la figura in alto nella pagina a fronte*).

Si può fare in modo di non deformare le bande e di indurre una corrente a

La giunzione *p-n* esiste fra una regione di un semiconduttore in cui gli elettroni carichi negativamente sono i portatori di corrente predominanti (materiale di tipo *n*) e un'altra regione in cui predominano le buche cariche positivamente (materiale di tipo *p*). Il campo elettrostatico indotto piega i margini della banda di valenza e della banda di conduzione alla giunzione, producendo una barriera di potenziale che impedisce agli elettroni e alle buche di attraversare la giunzione (*a sinistra*). Applicando una tensione esterna pari circa a E_g (procedimento che viene di solito chiamato polarizzazione della giunzione) si possono avere bande non piegate e ridurre le barriere di potenziale con conseguente flusso di corrente (*a destra*). Per l'arseniuro di gallio con appropriate impurezze, la maggior parte di corrente è portata attraverso la giunzione dagli elettroni che sono iniettati dalla sorgente esterna nella regione di tipo *p*. In quella regione la presenza degli elettroni è minoritaria ed essi non sono in equilibrio. Perciò si ricombinano con le buche, liberando in questo processo la loro energia sotto forma di luce.

fluire attraverso la giunzione *p-n* applicando al cristallo una tensione esterna in maniera che essa si opponga al campo indotto internamente. Questo processo si chiama « polarizzazione » del diodo. Quando la tensione di polarizzazione è applicata al diodo, gli elettroni sono indotti a fluire dal circuito esterno entro la parte *n* e a fuoruscire dalla parte *p* verso il circuito esterno. (Dire che gli elettroni fuoriescono dalla parte *p* verso il circuito esterno è equivalente a dire che le buche fluiscono dal circuito esterno verso la parte *p*). Quando la giunzione *p-n* è polarizzata sufficientemente, gli elettroni fluiscono attraverso la giunzione *p-n* dalla parte *n* della banda di con-

Quando gli elettroni sono iniettati attraverso una giunzione *p-n* a bassa corrente, si ricombinano spontaneamente con le buche e emettono luce incoerente in ogni direzione (*a*). Quando la corrente viene aumentata iniziano ad aver luogo emissioni stimolate vicino alla giunzione ove vi sono coppie eccitate costituite da una buca e da un elettrone che non hanno avuto tempo di ricombinarsi spontaneamente; la stimolazione avviene quando un fotone dovuto alla ricombinazione di una tale coppia colpisce un'altra coppia (*b*). Poiché la luce stimolata possiede la stessa fase e direzione della luce stimolante, una successione di tali eventi si risolve in una amplificazione. Quando il cristallo è fatto con una giunzione *p-n* piana e ha due facce di clivaggio perpendicolarmente al piano della giunzione, una parte della luce è riflessa dalle facce, che agiscono come specchi semiriflettenti (*c*). La luce è quindi intrappolata fra gli specchi e può contribuire a una ulteriore stimolazione e amplificazione. A una corrente sufficientemente elevata è maggiore la luce prodotta rispetto alla stimolazione di quella persa per assorbimento, diffusione, perdita all'esterno e così via; a quel punto la quantità di luce intrappolata inizia a crescere rapidamente e avviene il lasing, o azione laser. Queste illustrazioni mostrano come tali processi hanno luogo entro un laser a omostruttura.

Le eterogiunzioni possono essere fatte fra materiali semiconduttori diversi con tipo di conducibilità identica o differente. In queste illustrazioni i profili del potenziale della banda di valenza e della banda di conduzione sono mostrati per una eterogiunzione *p-p* e per una eterogiunzione *p-n* fra l'arseniuro di gallio (GaAs) e l'arseniuro di gallio e alluminio ($Al_xGa_{1-x}As$). Alla eterogiunzione *p-p* (*a sinistra*) la maggior separazione fra le bande nell'arseniuro di gallio e alluminio produce un gradino nella banda di conduzione che può agire come una barriera di potenziale per riflettere gli elettroni se sono iniettati dalla sinistra. La polarizzazione del diodo non ha effetto alcuno su questa barriera. Alla eterogiunzione *p-n*, se non vi è polarizzazione, le bande sono già piegate a impedire il flusso di corrente. Quando la polarizzazione è sufficiente per iniettare elettroni dall'arseniuro di gallio di tipo *n* all'arseniuro di gallio di tipo *p* (*a destra*), una barriera di potenziale residua rimane sulla banda di valenza alla eterogiunzione *p-n* e impedisce pertanto alle buche di attraversare l'eterogiunzione *p-n* dalla destra.

In questa illustrazione sono confrontati i principi del funzionamento di un laser a omostruttura (*a sinistra*), di un laser a eterostruttura singola (*al centro*) e di un laser a doppia eterostruttura. Nel laser a omostruttura gli elettroni (*frecce nere*) sono iniettati attraverso l'omogiunzione *p-n* a distanze differenti nella regione *p*. (L'iniezione delle buche non diviene importante finché non sono raggiunti livelli di corrente molto più alti). Nel dispositivo a omostruttura la luce (*frecce colorate*) penetra nel materiale adiacente al di fuori della regione ove ha luogo la maggior parte della stimolazione, causando una perdita di luce che altrimenti contribuirebbe alla stimolazione. Nel dispositivo a eterostruttura la luce non penetra significativamente attraverso l'eterogiunzione *p-p* e gli elettroni sono riflessi. Si ottiene come risultato un più alto grado di confinamento delle cariche e della luce. Nel dispositivo a doppia eterostruttura le buche, gli elettroni e la luce sono tutti confinati effettivamente nella regione molto sottile fra le eterogiunzioni. Gli elettroni sono riflessi dalla barriera di potenziale nella banda di conduzione alla eterogiunzione *p-p*, e le buche sono riflesse dalla barriera di potenziale nella banda di valenza alla eterogiunzione *p-n*. Il confinamento deriva dalla differenza nell'indice di rifrazione fra l'arseniuro di gallio e quello di gallio e alluminio.

DENSITÀ DI CORRENTE (AMPERE AL CENTIMETRO QUADRATO)

20 000 · 15 000 · 10 000 · 5000 · 0

0 · 1 · 2 · 3 · 4

LARGHEZZA DELLA REGIONE ATTIVA DEL LASING (MICRON)

In questo grafico sono confrontate la corrente di soglia per l'inizio dell'azione laser alla temperatura ambiente in un laser a eterostruttura singola (*curva in alto*) e in un laser a eterostruttura doppia (*curva in basso*) in funzione della larghezza della regione attiva per il lasing. Nel dispositivo a eterostruttura singola la corrente di soglia diminuisce con la larghezza della regione attiva per il lasing con l'avvicinarsi a due micron perché larghezze piú piccole corrispondono a piú piccoli volumi della regione attiva. Piú piccolo è il volume, minori sono gli elettroni necessari per raggiungere la densità elettronica richiesta per l'inizio del lasing. Al di sotto di circa due micron la polarizzazione necessaria a iniziare il lasing è sufficientemente grande per iniziare l'iniezione delle buche. Ciò significa che una parte della corrente è stata portata dalle buche che vanno nella regione *n*, ove non contribuiscono all'azione laser. Si ha come risultato un aumento della soglia del lasing. Questo effetto non si applica ai laser a eterostruttura doppia a causa della barriera di potenziale alla eterogiunzione *p-n*.

duzione verso la parte *p*: questo processo è chiamato iniezione degli elettroni. In molti semiconduttori le buche attraversano la giunzione *p-n* dalla parte *p* della banda di valenza verso la parte *n* della giunzione e questo processo è chiamato iniezione delle buche. Usualmente è predominante uno di questi processi. Per il semiconduttore che ci interessa, l'arseniuro di gallio, il processo predominante è l'iniezione di elettroni.

Una volta nella banda di conduzione della regione *p*, gli elettroni iniettati possiedono una energia maggiore dell'energia che loro compete normalmente in questa regione di una quantità approssimativamente uguale all'intervallo fra le bande. Gli elettroni rimangono nella banda di conduzione per un certo tempo (la loro « vita media ») e finalmente si ricombinano con le buche già presenti nella regione. Essi possono ricombinarsi in vari modi, ma nei diodi ad arseniuro di gallio una frazione sostanziale degli elettroni iniettati cede la propria energia attraverso l'emissione di un fotone di lunghezza d'onda determinata approssimativamente dall'intervallo di energia tra le bande dell'arseniuro di gallio. Il processo è chiamato elettroluminescenza, e un dispositivo che produce luce in questa maniera si chiama diodo a elettroluminescenza. Nell'arseniuro di gallio l'intervallo tra le bande è largo circa 1,4 elettronvolt. La luce emessa dai diodi luminescenti ad arseniuro di gallio possiede approssimativamente quella energia e perciò ha una lunghezza d'onda di circa 0,00009 centimetri. Un'onda elettromagnetica di questa lunghezza non è visibile ma è nella parte vicina al visibile della regione infrarossa dello spettro.

La luce cosí prodotta si dice dovuta a emissione spontanea perché le buche e gli elettroni si combinano spontaneamente dopo un certo intervallo di tempo medio senza essere influenzati dai fotoni degli altri eventi di ricombinazione. Come viene aumentata la corrente di iniezione comincia a divenire importante un processo chiamato emissione stimolata. In generale quando un elettrone in uno stato eccitato è colpito da un fotone con la stessa energia della sua energia di eccitazione, sarà stimolato a emettere quella energia sotto forma di un fotone di luce durante una transizione a un piú basso livello energetico; la luce stimolata avrà la stessa fase, la stessa lunghezza d'onda e la stessa direzione della luce stimolante. Gli elettroni iniettati nella regione *p* sono, naturalmente, elettroni eccitati, con una energia di eccitazione pari approssimativamente a E_g. Perciò, quando un elettrone iniettato che non ha ancora avuto la possibilità di ricombinarsi spontaneamente è colpito da un fotone emesso da un precedente evento di ricombinazione, può essere stimolato a emettere il suo fotone nella stessa direzione e con la stessa fase del fotone stimolante; in queste circostanze entrambi i fotoni avranno la stessa energia E_g.

Col crescere della corrente di iniezione aumenta la luce prodotta dalla emissione stimolata, poiché aumenta il numero di coppie buca-elettrone presenti vicino alla giunzione *p-n*; come risultato si ottiene una amplificazione, particolarmente lungo il piano di giunzione. Se il diodo è stato fatto con una giunzione *p-n* piana, esso può essere fabbricato in modo tale da avere due estremità piane del cristallo perpendicolari al piano della giunzione. Queste estremità piane (in realtà sono superfici di clivaggio del cristallo) agiscono come specchi parzialmente riflettenti e tendono a intrappolare parte della luce prodotta parallelamente al piano di giunzione riflettendola indietro. La struttura è ora un laser a diodo (*si veda la figura a pag. 53*).

L'introduzione degli specchi alimenta la luce stimolata entro il piano della giunzione e quindi favorisce l'amplificazione. Si dice che il diodo è « lasing » quando la corrente di iniezione diviene sufficientemente alta in modo che la luce nel cristallo, viaggiando avanti e indietro lungo il piano di giunzione, è amplificata abbastanza da controbilanciare perdite dovute all'assorbimento, alla diffusione, alla fuga all'esterno dello specchio e cosí via. Nel linguaggio dei ricercatori del settore, il guadagno ora eguaglia la perdita. Con un ulteriore aumento della corrente di iniezione le ulteriori coppie buca-elettrone si ricombinano principalmente per stimolazione. Si osserva un fascio di luce intensa emergere dalle estremità a specchio del cristallo; l'intero processo si dice di amplificazione della luce mediante emissione stimolata di radiazione (dall'inglese *light amplification by stimulated emission of radiation*, deriva l'acronimo « laser »).

I semplici laser a diodo del tipo descritto sopra sono chiamati laser a omostruttura poiché sono fatti di un singolo materiale semiconduttore. Anche se sono costruiti con il migliore materiale disponibile (arseniuro di gallio), il lasing a temperatura ambiente richiede alte densità di corrente (da 25 000 a 100 000 ampere al centimetro quadrato). In altre parole, se l'area del piano di giunzione del diodo a laser fosse un centimetro quadrato, sarebbe necessaria nei migliori dei casi una corrente di 25 000 ampere. In realtà, naturalmente, le aree sono molto piú piccole, di solito di 0,0005 centimetri, e la corrente corrispondente è di 12,5 ampere. L'uso del concetto di « densità di corrente » ci mette in grado di confrontare diodi con aree differenti. La densità di corrente minima necessaria a innescare l'azione di lasing in un diodo è definita densità di corrente di soglia. È a causa della loro alta densità di corrente di soglia che tali laser sono sempre fatti funzionare a impulsi alla temperatura ambiente allo scopo di mantenere sufficientemente bassa la dissipazione di potenza per impedire il surriscaldamento. Il funzionamento continuo, o non impulsato, è possibile solo a temperature molto piú basse (di solito 77 gradi Kelvin, punto di ebollizione dell'azoto liquido), poiché la corrente di soglia del lasing è molto piú bassa a temperature piú basse.

Negli anni recenti la temperatura massima alla quale possono essere fatti funzionare con continuità i laser a omostruttura è stata di 205 gradi Kelvin. Questa temperatura fu raggiunta nel 1967 da J.C. Dyment e L.A. D'Asaro nei Laboratori Bell in un esperimento dove il laser era montato su un diamante per permettere una efficiente rimozione del calore. Da quel lavoro si riconobbe che, allo scopo di far funzionare un laser a diodo semiconduttore con corrente continua alla temperatura ambiente, non soltanto era necessario un efficiente pozzo di calore ma che anche si doveva realizzare una grande riduzione nella densità di corrente di soglia alla temperatura ambiente.

Retrospettivamente si possono tenere in giusto conto le ragioni per credere che poteva essere ottenuta una considerevole riduzione nella densità di corrente di soglia, sulla base del se-

guente ragionamento semplificato. Nella regione vicina al piano della giunzione *p-n*, dove nel laser a diodo semiconduttore ha luogo l'emissione stimolata, deve essere raggiunta una densità critica di luce e di elettroni affinché la amplificazione dovuta all'emissione stimolata superi le perdite. Nei diodi laser ad arseniuro di gallio, comunque, la distanza che gli elettroni possono percorrere allontanandosi dalla giunzione *p-n* dopo l'iniezione aumenta con la temperatura. Alla temperatura ambiente la distanza media di diffusione dell'elettrone dalla giunzione *p-n* dopo l'iniezione può essere dell'ordine di diversi micron. Pertanto risulta incontrollata l'ampiezza della regione che contribuisce al lasing; essa dipende non solo dal semiconduttore usato ma anche dalla temperatura e dalle impurità presenti. Ciò significa che il volume di materiale semiconduttore che deve raggiungere la densità elettronica richiesta non è un parametro che si sia in grado di modificare in un laser a omostruttura. Chiaramente se si potesse ridurre questo volume, si potrebbe ottenere una piú bassa densità di corrente di soglia.

Inoltre, gran parte della luce non è ben confinata in un tal laser; la luce tende a sparpagliarsi in regioni inattive del cristallo. Ciò riduce la densità di fotoni disponibile per contribuire all'amplificazione poiché la luce occupa un volume piú grande e poiché una parte di essa si è diffusa esternamente in regioni ove vi sono pochi elettroni eccitati. La luce è assorbita piuttosto che amplificata poiché la sua energia è approssimativamente la stessa della energia di separazione tra le bande dell'arseniuro di gallio. Perciò, in prima approssimazione, l'alta densità di corrente di soglia occorrente affinché avvenga il lasing alla temperatura ambiente nei laser a omostruttura è richiesta non solo dalla necessità di mantenere una data densità elettronica in un grande volume della regione *p* in cui le cariche si diffondono ma anche dalla grande perdita di luce e dalla ridotta densità fotonica risultante dallo sparpagliamento della luce laser.

L'ottenimento di un laser a soglia bassa fatto per funzionare con continuità alla temperatura ambiente deriva direttamente dall'uso di strutture modificate che costringono sia la luce sia gli elettroni in una piccola regione attiva. Si ottenne ciò utilizzando cristalli che avevano una regione di arseniuro di gallio puro e una regione ove alcuni degli atomi di gallio nei cristalli di arseniuro erano stati rimpiazzati da atomi di alluminio. L'ultima regione è costituita dal semiconduttore arseniuro di gallio e alluminio, e la giunzione dei due materiali si chiama eterogiunzione. Il simbolo chimico dell'arseniuro di gallio è GaAs e quello dell'arseniuro di gallio e alluminio è $Al_xGa_{1-x}As$, ove x è minore o uguale a 1. Il simbolismo indica la sostituzione uno a uno degli atomi di gallio da parte degli atomi di alluminio nel reticolo. Se, per esempio, metà degli atomi di gallio sono sostituiti da atomi di alluminio, la formula del semiconduttore risultante è $Al_{0,5}Ga_{0,5}As$.

Nell'arseniuro di gallio e alluminio la distanza tra le bande è piú grande di quella corrispondente nell'arseniuro di gallio; l'ammontare di questo aumento è esso stesso crescente con la frazione di atomi di gallio che sono sostituiti da atomi di alluminio. L'aumento della distanza ha effetti profondi non solo sul comportamento delle buche e degli elettroni quando essi incontrano queste eterogiunzioni ma anche nella luce elettroluminescente generata. Le buche e gli elettroni sono influenzati perché la presenza della eterogiunzione si risolve nel brusco cambiamento della distanza fra le bande e quindi della forma del fondo della banda di conduzione o della sommità della banda di valenza nelle sue vicinanze (*si veda la figura in alto a pag. 54*).

L'elettroluminescenza è influenzata perché l'arseniuro di gallio e alluminio è trasparente alla luce che ha l'energia della distanza fra le bande dell'arseniuro di gallio (quale risultato della piú ampia banda proibita dell'arseniuro di gallio e alluminio) e perché vi è una diminuzione dell'indice di rifrazione di mano in mano che aumenta E_g all'aumentare di x. Nel passato, quando vennero usati altri materiali per la preparazione di eterostrutture, queste proprietà non potevano essere utilizzate perché un'altra proprietà delle eterogiunzioni è quella che esse tendono a intrappolare buche ed elettroni in maniera tale da impedire che la ricombinazione avvenga con produzione di luce utile. Questo non accade con eterogiunzione fra arseniuro di gallio e arseniuro di gallio e alluminio. Pensiamo che ciò sia dovuto al fatto che i due composti hanno la stessa struttura cristallina e quasi esattamente la stessa spaziatura fra gli atomi del reticolo cristallino.

Come contribuiscono tutti questi fattori nel raggiungimento di una soglia ridotta del lasing nei laser a diodo? Per fare tali laser bisogna prima di tutto essere capaci di fare strutture cristalline con una o piú eterogiunzioni tra arseniuro di gallio e arseniuro di gallio e alluminio. Tali strutture sono generalmente fatte attraverso la crescita di entrambi i composti su porzioni di cristalli singoli di arseniuro di gallio. Allo stato attuale la crescita di questi strati avviene sui cristalli di arseniuro di gallio ad alte temperature da soluzioni liquide contenenti una miscela di gallio, alluminio e arsenico.

Rappresentazione schematica dell'apparato usato per far crescere strutture pluristratificate che servono come laser a diodo. La « barca » di grafite contiene quattro pozzi riempiti con soluzioni di gallio, alluminio, arsenico e un opportuno elemento di impurezza. Un wafer costituito da una lamina di un singolo cristallo di arseniuro di gallio viene montato su un contenitore di grafite scorrevole e quindi portato sotto ciascuna soluzione. Quando è usato un opportuno ciclo di raffreddamento, ciascuna soluzione fa crescere strati successivi. Da uno a quattro strati di arseniuro di gallio o di arseniuro di gallio e alluminio, ciascuno con un proprio complemento di impurezza, può essere fatto crescere su wafer. Le misure del wafer sono di quattro millimetri di larghezza per 10 millimetri di lunghezza e per mezzo millimetro di spessore. Le temperature variano fra 800 e 1000 °C.

I laser sono fabbricati dai wafer semiconduttori e pluristratificati dapprima ricoprendo il substrato del wafer fino a uno spessore totale che varia dai 75 ai 125 micron. Contatti elettrici di metallo molto sottili sono quindi applicati dirigendo fasci atomici di elementi appropriati sulle superfici in un sistema ad alto vuoto. Il cristallo è quindi tagliato in barre larghe approssimativamente mezzo millimetro, e queste a loro volta tagliate a fettine larghe circa 0,1 mm mediante un sottile filo metallico.

L'uso piú semplice che può essere fatto della natura ideale della eterogiunzione arseniuro di gallio-arseniuro di gallio e alluminio per la riduzione della soglia laser è quella del laser a eterostruttura singola. Questa struttura del laser è fatta attraverso la crescita di uno strato di arseniuro di gallio e alluminio su un substrato di arseniuro di gallio di tipo *n* e simultaneamente diffondendo zinco entro il substrato a breve distanza dalla soluzione in crescita. Lo zinco è un'impurità accettrice nell'arseniuro di gallio, e una quantità sufficiente di esso può convertire materiale di tipo *n* in materiale di tipo *p*. Lo zinco ha anche la proprietà che i suoi atomi possono, a temperature sufficientemente alte, muoversi prontamente entro il reticolo di arseniuro di gallio da una sorgente esterna, sia esso sotto forma di vapore, liquido o solido. Il risultato è allora una piastrina consistente principalmente di un substrato di tipo *n* ma con una sottile regione di tipo *p* causata dallo zinco diffuso. Sulla superficie della regione di tipo *p* vi è uno strato di arseniuro di gallio e alluminio («drogato» con zinco) di tipo *p*. La struttura possiede perciò una giunzione *p-n* nell'arseniuro di gallio (una omogiunzione) separata per mezzo di una piccola quantità di arseniuro di gallio di tipo *p* da uno strato di arseniuro di gallio e alluminio di tipo *p*. Ci riferiamo all'ultimo contorno come all'eterogiunzione *p-p*. Il piano di giunzione *p-n* e il piano di eterogiunzione *p-p* sono paralleli l'uno all'altro. La struttura ha una singola eterogiunzione, da cui il nome di laser a eterogiunzione singola.

Nel costruire gli attuali laser dalla piastrina il substrato è assottigliato sino a uno spessore totale di circa 100 micron per mezzo della erosione con un abrasivo. Contatti elettrici (strati sottili di metallo) sono applicati mediante deposizione sotto vuoto sulla parte *n* e sulla parte *p* del dischetto, e la composizione risultante è tagliata in piastrine rettangolari di 100 per 400 micron (*si veda la figura in alto*). Questi sono i laser. Quando un frammento è sufficientemente polarizzato affinché avvenga la iniezione di elettroni attraverso la giunzione *p-n*, gli elettroni iniettati vanno verso la regione di tipo *p* dell'arseniuro di gallio ma sono fermati (riflessi) da una barriera di potenziale formata da una flessione nella banda di conduzione ove vi è la eterogiunzione. Gli elettroni non sono perciò liberi di diffondersi in un volume piú grande, come accade con i laser a omostruttura.

Inoltre, quale risultato della diminuzione dell'indice di rifrazione nella eterogiunzione vi è una piccola penetrazione della luce al di là del confine della eterogiunzione. La corrente di soglia del lasing è quindi ridotta al di sotto di quella dei laser a omostruttura poiché sia gli elettroni che la luce nel laser a eterostruttura singola possono essere confinati in un volume piú piccolo. La soglia del lasing è ora dipendente dalla distanza fra la giunzione *p-n* e la eterogiunzione *p-p* poiché una distanza piú piccola significa che gli elettroni occupano un volume piú piccolo alla densità richiesta per il lasing (*si veda l'illustrazione a pagina 56*). Con il laser a eterostruttura singola una parte di luce è ancora perduta a causa della penetrazione attraverso la giunzione *p-n* nell'arseniuro di gallio di tipo *p*. Inoltre, quando la tensione necessaria a causare il flusso di corrente viene aumentata sufficientemente, alcune buche sono iniettate attraverso la giunzione *p-n* nell'arseniuro di gallio di tipo *n*. L'iniezione di buche accade principalmente nei laser a eterostruttura singola con una regione attiva molto stretta. Poiché questa parte di corrente non è di utilità per il lasing, la corrente di soglia aumenta. Anche con tali difetti sono stati costruiti laser a eterostruttura singola con una bassa soglia a temperatura ambiente, circa 8000 ampere al centimetro quadrato.

Una ulteriore considerevole riduzione nella soglia di corrente si ottiene se alla giunzione *p-n* viene collocata una eterogiunzione addizionale. Allo scopo di realizzare una tale eterogiunzione *p-n* sul substrato di arseniuro di gallio bisogna che cresca una struttura a tre strati. In realtà gli strati cresciuti sono quattro (*si veda l'illustra-*

zione in basso a pag. 51). Il quarto strato (di arseniuro di gallio) serve per fare contatto elettrico.

La parte importante di questa struttura stratificata è un sandwich di arseniuro di gallio e alluminio di tipo *n*, di arseniuro di gallio di tipo *p* e di arseniuro di gallio e alluminio di tipo *p*; il sandwich comprende i primi tre strati e il substrato. Una eterogiunzione *n-p* esiste fra i primi due strati e una eterogiunzione *p-p* fra i secondi due. La regione di arseniuro di gallio di tipo *p* fra queste due eterogiunzioni definisce la regione attiva del lasing; tale struttura si chiama eterostruttura doppia.

Come prima, sono applicati i contatti e il dischetto con gli strati è spezzato in piccoli frammenti laser rettangolari. Quando questi sono sufficientemente polarizzati affinché gli elettroni siano iniettati nel secondo strato, sul bordo della banda di valenza rimane un gradino residuo. Come prima, gli elettroni sono riflessi dalla eterogiunzione *p-p*, ma ora le buche sono anche riflesse dalla eterogiunzione *p-n*. Il risultato di questa seconda riflessione è che a nessuna tensione prossima alla tensione necessaria per ottenere il lasing vi è un guadagno nelle iniezioni di buche col conseguente aumento della corrente di soglia. Pertanto è possibile costruire laser con regioni attive molto piú strette di quanto fosse possibile con il laser a eterostruttura singola. Vi sono adesso dei salti dell'indice di rifrazione a entrambe le eterogiunzioni *p-p* e *p-n* e la penetrazione della luce all'esterno della regione attiva è ristretta in entrambe le parti. Abbiamo cosí una struttura che per larghezze della regione attiva dell'ordine di 0,5 micron possiede una quasi completa limitazione della luce e della ricombinazione delle cariche. Ciò significa che è possibile avere una regione attiva molto piccola in cui sono confinati buche, elettroni e luce. Perciò allo scopo di ottenere una sufficiente densità di cariche e di fotoni per l'inizio del lasing è necessaria una corrente molto piú bassa di quella richiesta dai laser a omostruttura o da quelli a eterostruttura singola. Con i laser a eterostruttura doppia sono state ottenute densità di corrente di soglia per l'azione laser dell'ordine di circa 1000 ampere al centimetro quadrato, e si crede che da ulteriori ricerche si otterranno soglie piú basse.

Sebbene il raggiungimento di tali densità di corrente di soglia relativamente bassa avvicini notevolmente il raggiungimento del funzionamento continuo dei laser a diodo, rimane ancora da risolvere un problema piú grande. Durante il funzionamento a correnti di circa 1000-3000 ampere al centimetro quadrato molto calore deve essere rimosso dal laser. Poiché la maggior parte del calore è generato nella regione attiva, esso deve essere rimosso attraverso almeno uno strato di arseniuro di gallio e alluminio. Tali composti ternari hanno una conducibilità termica molto piú bassa di quella dei composti binari quali l'arseniuro di gallio. Perciò qui il problema è piú complicato di quanto fosse con i laser a omostruttura. Per superare questa difficoltà almeno uno dei due strati di arseniuro di gallio e alluminio deve essere molto sottile. Le stime indicano che, rendendo uno di questi strati spesso circa un micron, sarebbe possibile il funzionamento continuo dei laser a bassa soglia se il laser viene inserito in un efficiente pozzo di calore.

Si raggiunse un efficiente pozzo di calore amalgamando il contatto metallico sulla superficie di tipo *p* del laser a un diamante rivestito di stagno. In questa maniera sono state ottenute nei nostri laboratori le prime unità che funzionavano a temperatura ambiente o superiore. La struttura ottenuta in questo modo è una semplice configurazione a piani paralleli. Recentemente sono stati ottenuti pozzi di calore piú efficienti con una configurazione a « strisce » (*si veda l'illustrazione in basso*). Con questa configurazione soltanto una piccola striscia della regione attiva è percorsa da corrente. Viene prodotto minor calore perché la parte attiva del dispositivo è piú piccola, e il calore che viene prodotto è dissipato piú rapidamente perché l'eliminazione del calore avviene tanto lateralmente quanto verticalmente nella struttura a strisce. È già stato collaudato un notevole numero di tali unità. L'utilizzazione continua di tali laser a strisce è stata ottenuta sia con pozzi di calore a diamante che a rame; i laser sono stati fatti funzionare con continuità a temperature di circa 100 gradi centigradi con pozzi di calore a diamante e con modificazioni della geometria a strisce.

Abbiamo descritto l'evoluzione di una nuova classe di laser, i laser a eterostruttura, che finora sono stati realizzati con potenze di uscita di decine di milliwatt mentre funzionavano con continuità a lunghezze d'onda comprese fra 0,00009 e (con lievi modificazioni della struttura) 0,000077 centimetri. Durante il funzionamento continuo di parecchi laser di questo tipo sono già state raggiunte efficienze totali di circa il 10 per cento nella conversione di energia elettrica in energia ottica. Queste cifre saranno probabilmente migliorate con ulteriori studi. Sebbene i dispositivi siano ancora allo stadio di ricerca, essi hanno notevoli prospettive quali generatori di onde portanti piccoli, a buon mercato, efficienti e robusti per sistemi di comunicazioni di massa di tipo ottico e per parecchie altre applicazioni.

Il disegno di questo laser a doppia eterostruttura è caratterizzato da una geometria a strisce. Lo strato di ossido isolante restringe la corrente a una stretta striscia lungo la regione attiva, permettendo al calore di uscire rapidamente dalla regione del lasing.

Fotoni "pesanti"

Fotoni con energia miliardi di volte superiore a quella dei fotoni della luce visibile hanno proprietà che si pensavano proprie dei soli adroni

di Frederick V. Murphy e David E. Yount

Lo studio del comportamento e della natura della luce ha tenuto gli scienziati occupati per secoli e ha dato origine a innumerevoli dispute. Poiché la luce presenta caratteristiche corrispondenti sia a onde sia a particelle, essa ha fornito ai ricercatori un primo esempio dei paradossi che la natura ha in serbo quando è indagata al suo livello piú fondamentale – paradossi che portarono infine allo sviluppo della meccanica quantistica. Nei tre decenni che seguirono questa conquista scientifica, il comportamento dei quanti di luce, o fotoni, divenne cosí ben compreso e cosí prevedibile da far pensare che altre sorprese fossero impensabili.

Ma le sorprese ci furono. Esperimenti effettuati negli ultimi anni con acceleratori di altissima energia hanno dimostrato che il fotone può avere proprietà attribuite una volta solamente a particelle dotate di massa come il protone e il neutrone, i pesanti e stabili costituenti del nucleo atomico. Simili particelle, che esercitano fra di loro la forza « forte » che mantiene assieme il nucleo (detta anche « interazione forte »), sono chiamate adroni. Si è trovato che un fotone avente un miliardo di volte l'energia di un fotone della luce visibile si comporta come gli adroni se gli viene consentito di interagire con essi. La scoperta che il fotone ha queste caratteristiche adroniche per energie molto elevate, benché inaspettata, è stata adesso incorporata in un nuovo gruppo di teorie che rendono accettabile il fotone adronico.

Prima del XVII secolo l'opinione piú diffusa era che la luce consistesse di un fascio di particelle emesse in direzione radiale da una sorgente, quale il Sole o la fiamma di una candela. Alcuni materiali erano trasparenti a queste particelle; altri erano opachi o traslucidi e tendevano a rifletterle o ad assorbirle in varia misura. Nel 1678 Christiaan Huygens dimostrò che se la luce fosse costituita da onde susseguentesi invece che da particelle indipendenti, le leggi della riflessione e della rifrazione potevano essere spiegate in modo semplice. All'inizio del XIX secolo Thomas Young e Augustin Jean Fresnel fecero indipendentemente esperimenti sull'interferenza e la diffrazione che si accordavano con l'ipotesi ondulatoria e che fornivano una misura della lunghezza d'onda della luce.

Nella seconda metà del XIX secolo James Clerk Maxwell riuscí a sintetizzare i fenomeni elettrici e magnetici in quattro equazioni che sono ancora alla base del funzionamento di tutti i dispositivi elettromagnetici. Le equazioni di Maxwell predicono che un circuito elettrico oscillante emetterà onde elettromagnetiche la cui velocità può essere calcolata da misure puramente elettromagnetiche. Quando la velocità calcolata risultò essere esattamente quella determinata sperimentalmente per la luce, i fisici si convinsero che la luce consiste non solamente di onde ma proprio di quelle particolari onde elettromagnetiche previste dalle equazioni dell'elettromagnetismo. Perciò le equazioni di Maxwell portarono a pensare all'esistenza di un ampio spettro elettromagnetico che si estendesse da ambo le parti dello spettro visibile e includesse le onde radio scoperte da Heinrich Hertz nel 1886.

La natura corpuscolare della luce fu reintrodotta da Albert Einstein nel 1905 per spiegare perché la luce che colpisce un metallo estrae elettroni dalla superficie del metallo. L'energia della luce, invece di essere distribuita in un'onda estesa, è concentrata in piccoli pacchetti chiamati fotoni. Quando un singolo fotone colpisce un elettrone sulla superficie del metallo, il fotone è completamente assorbito e imparte tutta l'energia all'elettrone. Curiosamente, Einstein predisse, e Robert A. Millikan dimostrò sperimentalmente, che l'energia del fotone, e quindi quella dell'elettrone emesso, dipende solamente dalla frequenza della luce, come se ciascun fotone fosse simultaneamente anche un'onda. Un altro esempio del carattere corpuscolare della luce fu scoperto nel 1921 da Arthur H. Compton, che osservò la diffusione dei fotoni nell'urto con elettroni: i fotoni si comportano come corpi materiali aventi energia e quantità di moto che vengono conservate negli urti. La dualità onde-particelle è risolta nell'elettrodi-

BERILLIO (9)
CARBONIO (12)
ALLUMINIO (27)
RAME (64)
ARGENTO (108)
TANTALIO (181)
PIOMBO (207)
URANIO (238)
PRODUZIONE DEI RHO
MOMENTO TRASVERSALE
MASSA (MeV)

La produzione e la diffusione di mesoni rho sono studiate bombardando bersagli aventi masse nucleari diverse con fotoni di alta energia. La vita media dei mesoni rho è cosí breve (circa 10^{-23}s) che di solito decadono in due pioni prima che possano colpire un altro nucleo e venirne diffusi. Nella figura il tasso di produzione dei mesoni rho, dedotto dai loro prodotti di decadimento, è posto in un grafico per otto diversi materiali del bersaglio, che vanno dal berillio, con soli otto nucleoni (protoni o neutroni), all'uranio, che ne ha 238. Il tasso della produzione di rho per nucleone del bersaglio è posto in ordinata in unità arbitrarie. La massa dei rho è rappresentata lungo l'asse delle ascisse; le unità sono milioni di elettronvolt (MeV). Il terzo asse rappresenta il quadrato della quantità di moto trasversale impressa al nucleo bersaglio; le unità sono $(\text{GeV}/c)^2 \times 10^{-3}$ (1 GeV = 10^9 eV; c è la velocità della luce). Per un dato valore della quantità di moto trasversale trasmessa, l'area sotto la curva a campana che rappresenta la massa dei rho aumenta dapprima con il numero dei nucleoni del bersaglio. In prossimità del rame (64 nucleoni) la produzione dei rho per nucleone comincia a stabilizzarsi oppure a diminuire coll'aumento del numero dei nucleoni. Nei nuclei piú pesanti vengono prodotti sempre piú rho, ma essi vengono anche assorbiti in maniera crescente. I dati sono stati forniti dall'elettrosincrotrone di Amburgo (DESY). Un gruppo di ricercatori del MIT ha collaborato all'esperimento con il gruppo del DESY.

L'allineamento dell'acceleratore lineare di Stanford (SLAC) viene compiuto con l'aiuto di un fascio laser e di bersagli di Fresnel. I bersagli sono situati ogni 12,2 metri in una guida di luce che è collegata rigidamente a/e sotto il tubo da 2,5 centimetri di diametro che trasporta il fascio di elettroni accelerati. Quando i bersagli sono colpiti da un fascio laser, essi producono una caratteristica figura di diffrazione se la sezione del tubo è esattamente allineata. Tre martinetti che servono per allineare il tubo sono visibili nella fotografia all'estrema sinistra, in basso a sinistra e in basso a destra.

namica quantistica, che è forse la teoria più elegante e di maggior successo della fisica. Essa prevede le interazioni elettromagnetiche dei fotoni con così grande precisione e su di una tale gamma di energie (per un rapporto di 10^{18} a uno) da rendere possibili molti esperimenti rigorosi. Poiché i progressi della fisica si basano spesso sul fallimento delle teorie, i continui successi dell'elettrodinamica quantistica sono stati accolti quasi con dispiacere.

Fino a poco tempo fa l'interazione dei fotoni con la materia si pensava fosse interamente elettromagnetica, cioè sembrava interessare solamente le cariche elettriche e i campi magnetici delle particelle bersaglio e non le loro possibili interazioni adroniche, o forti. Per di piú, i fotoni fungono da mediatori di tutte le interazioni elettromagnetiche: due elettroni si respingono palleggiandosi dei fotoni avanti e indietro. I fotoni scambiati esistono per tempi estremamente brevi e non possono essere rivelati direttamente. Per questa ragione essi sono detti « virtuali » *(si veda « a » nella figura della pagina a fronte).*

Quando un elettrone viene diffuso da un altro elettrone e da un'altra particella carica, esso è accelerato bruscamente e può irraggiare altri fotoni, che non vengono riassorbiti. Questa radiazione, detta *bremsstrahlung* (radiazione di frenamento), è analoga alla emissione di onde radio da parte di un circuito elettrico oscillante. Come le onde radio, i fotoni di *bremsstrahlung* possono essere rivelati e sono perciò detti « reali » *(si veda « b » nella figura della pagina a fronte).* I fisici applicano la loro conoscenza delle interazioni elettromagnetiche alla produzione di grandi quantità di fotoni reali o virtuali di energia nota in condizioni controllate, cioè alla produzione di fasci di fotoni. La disponibilità di tali fasci controllati di fotoni di alta energia fu la chiave della scoperta del carattere adronico dei fotoni.

Quando un elettrone irraggia un fotone reale di alta energia durante una collisione con un nucleo atomico pesante, l'energia trasformata in fotoni virtuali è di solito piccola e il nucleo rimane quasi fermo. L'energia del fotone reale è quindi eguale semplicemente alla differenza fra l'energia dell'elettrone prima e dopo la collisione. Questa energia può essere misurata facilmente deviando gli elettroni, incidente e diffuso, in forti campi magnetici. In questo modo si produce un fascio di fotoni reali « etichettati » in cui ogni fotone del fascio ha una data energia.

I fisici si interessano anche dei processi che coinvolgono fotoni virtuali e nuclei. Se non vengono irraggiati fotoni reali durante una collisione elettrone-nucleo, l'energia del fotone virtuale è ancora eguale alla differenza di energia dell'elettrone prima e dopo la collisione. Quando una parte dell'energia del fotone virtuale viene impiegata per spezzare il nucleo o per creare nuove particelle, per esempio pioni, il processo è detto anelastico. La diffusione anelastica degli elettroni è quindi analoga all'etichettatura dei fotoni reali e rende possibili esperimenti con fotoni virtuali non rivelati, simili a quelli condotti con fotoni reali.

Un secondo metodo per produrre un fascio di fotoni reali di alta energia è quello di diffondere la luce visibile di 180° con un fascio di elettroni di alta energia, esattamente come una palla da tennis può essere « diffusa » di 180° da una racchetta in movimento. Si tratta ancora di una diffusione alla Compton, anche se in questo caso gli elettroni si muovono quasi alla velocità della luce quando avviene l'urto frontale. Come nell'esperimento originale di Compton i fotoni diffusi si comportano come corpi materiali aventi un'energia e una quantità di moto ben definite che possono essere calcolate in base alle condizioni sperimentali. E mentre la palla da tennis colpita dalla racchetta può aumentare la sua energia di dieci volte, l'energia del fotone diffuso può aumentare di un fattore pari a 10 miliardi. Un fascio di questo tipo, avente un laser impulsato di alta energia come fonte primaria di luce, dimostra la corrispondenza diretta esistente fra fotoni appartenenti a regioni molto distanti dello spettro elettromagnetico.

Un altro metodo usato per produrre fotoni reali o virtuali è l'annichilazione fra particelle e antiparticelle, che avviene per esempio quando un elettrone urta la sua antiparticella, il positone. Quando un fascio di positoni colpisce un bersaglio, un positone e un elettrone possono interagire fino a produrre due fotoni reali: due quanti di energia elettromagnetica senza massa *(si veda « c » nella figura qui sotto).* Occorrono due fotoni, invece di uno, per la conservazione dell'energia e della quantità di moto. (Questo processo è l'inverso della « produzione di coppie » in cui un solo fotone reale si materializza letteralmente creando una coppia elettrone-positone. È necessario un secondo fotone anche nella produzione di coppie: un fotone virtuale che conserva l'energia e la quantità di moto trasferendo queste quantità a un nucleo vicino). L'annichilazione fra elettroni e positoni può anche dare un solo fotone virtuale che, se dotato di energia sufficiente, può materializzarsi direttamente in un'altra coppia particel-

a — e⁻ e⁻ FOTONE VIRTUALE e⁻ e⁻ — DIFFUSIONE

b — e⁻ e⁻ FOTONE VIRTUALE FOTONE REALE p p — RADIAZIONE

c — e⁺ FOTONE REALE ELETTRONE VIRTUALE e⁻ FOTONE REALE e⁻ — ANNICHILAZIONE REALE

d — FOTONE REALE e⁺ ELETTRONE VIRTUALE e⁻ e⁻ FOTONE VIRTUALE p p — PRODUZIONE DI COPPIE

e — e⁺ FOTONE VIRTUALE π⁺ e⁻ π⁻ — ANNICHILAZIONE VIRTUALE

f — ACCOPPIAMENTO RHO-FOTONE e⁺ FOTONE VIRTUALE π⁺ RHO e⁻ π⁻ — ANNICHILAZIONE VIRTUALE MEDIATA DA UN RHO

Le interazioni fra particelle sono rappresentate di solito mediante diagrammi di Feynman. Le particelle « virtuali » non rivelabili sono necessarie per spiegare in quale modo l'energia e la quantità di moto sono scambiate durante una reazione. Secondo il principio di indeterminazione, i fotoni virtuali possono aver una massa diversa da zero, a differenza dei fotoni reali, che sono privi di massa. Nella diffusione elettrone-elettrone *(a)*, viene scambiato un fotone virtuale. Quando un elettrone viene diffuso da un protone *(b)*, un fotone reale può essere emesso e usato in esperimenti che richiedano fasci di fotoni. Tali fasci possono essere anche creati dai fotoni prodotti dall'annichilazione *(c)* di un elettrone (e^-) e di un positone (e^+). L'inverso dell'annichilazione *(d)* produce una coppia elettrone-positone. L'annichilazione elettrone-positone può anche dare un solo fotone virtuale *(e)*, che può quindi produrre una coppia di pioni (π^+, π^-). La materializzazione di un fotone virtuale in due pioni è aumentata quando il fotone virtuale si materializza dapprima in un mesone vettoriale rho.

la-antiparticella quale una coppia protone-antiprotone o una coppia di pioni positivo e negativo (*si veda « e » nella figura della pagina precedente*).

Qui c'è una contraddizione apparente. Nell'annichilazione elettrone-positone un fotone reale non è sufficiente per la conservazione dell'energia e della quantità di moto. Com'è possibile, allora, l'annichilazione in un solo fotone virtuale? È possibile perché il fotone virtuale, a differenza del fotone reale, non è un prodotto finale e non può essere rivelato; perciò non c'è bisogno che sia privo di massa. Poiché il fotone virtuale esiste solamente per un tempo molto breve, secondo il principio di indeterminazione di Heisenberg esiste una incertezza molto grande sulla sua energia. Entro questa indeterminazione la quantità di moto e l'energia non sono necessariamente conservate durante il breve tempo di interazione, benché siano conservate per le particelle iniziali e finali che vivono abbastanza a lungo per essere rivelate. Dire che l'energia e la quantità di moto non si conservano equivale a dire che il fotone virtuale ha massa diversa da zero: piú breve è la sua vita, piú ampia la gamma di masse che una particella virtuale può assumere. Un fotone reale può avere anch'esso una massa diversa da zero durante il breve periodo in cui interagisce, ma questa momentanea deviazione non può essere misurata direttamente anche in questo caso. Perciò i fotoni reali vengono sempre creati a coppie nell'annichilazione di un elettrone e un positone, in quanto possono essere rivelati. Se, d'altra parte, il prodotto dell'annichilazione è un fotone virtuale, è sufficiente un fotone poiché esso diventa qualcosa d'altro prima che si possa osservare la deviazione della sua massa.

Il principio di indeterminazione ha un ruolo ancora piú importante nel ragionamento che lega i fotoni agli adroni. La capacità dei fotoni di assumere una vasta gamma di masse durante le interazioni di alta energia rende possibile il fatto che essi assumano la massa e altre caratteristiche di certi adroni detti mesoni vettoriali. Le particelle elementari possono essere classificate assegnando a ciascuna di esse un insieme di numeri detti numeri quantici (per esempio i numeri che rappresentano la carica elettrica e il momento angolare di spin). Noti i numeri quantici di alcune particelle, si possono prevedere le interazioni possibili fra esse. Ciò semplifica un altrimenti indecifrabile intrico di interazioni.

I mesoni neutri vettoriali di « stranezza » zero (una quantità intrinseca di tipo particolare) hanno gli stessi numeri quantici dei fotoni. Anche nel loro stato libero, reale, i mesoni vettoriali non sono privi di massa e, poiché interagiscono fortemente con altri mesoni e con nucleoni, essi devono essere classificati come adroni. La meccanica quantistica insegna che se due particelle differiscono solamente per la massa, allora, qualora abbiano la stessa massa apparente, devono partecipare alle stesse interazioni. Ciò facilita grandemente le interazioni fotone-adrone che si possono dividere in due fasi: dapprima il fotone si materializza in un mesone vettoriale nelle vicinanze dell'adrone interagente e quindi il mesone vettoriale interagisce fortemente con l'adrone (*si veda « f » nella figura della pagina precedente*). Per energie molto grandi questo processo in due fasi è piú probabile di un'interazione singola diretta poiché la materializzazione dei fotoni in mesoni vettoriali avviene piú prontamente quando l'energia e l'indeterminazione della massa virtuale sono grandi, e poiché la susseguente interazione mesone-adrone è intrinsecamente 200 volte piú probabile dell'interazione diretta fotone-adrone.

Un modello particolarmente utile di interazioni fotone-adrone, chiamato metodo a dominanza vettoriale, è stato introdotto da J.J. Sakurai nel 1960 e fu in seguito sviluppato da un buon numero di altri ricercatori. Il modello è basato essenzialmente su tre relazioni di equivalenza. La prima dice che durante un'interazione fotone-adrone il fotone incidente è equivalente a una combinazione di mesoni vettoriali. La seconda relazione dice che ogni fotone derivante da un'interazione di questo tipo è una combinazione di mesoni vettoriali. La terza relazione è un venerando principio, scoperto dapprima nell'ottica ordinaria, che mette in rapporto l'assorbimento della luce con la diffusione della stessa per diffrazio-

La massa del mesone rho viene misurata trovando l'energia del fotone virtuale che dà il piú alto numero di coppie π^+, π^-, particelle che vengono prodotte nel decadimento del mesone rho. L'energia del fotone virtuale è a sua volta la somma delle energie dell'elettrone e del positone che si sono annichilati per produrlo. I dati sono stati ottenuti a Orsay, in Francia, facendo urtare un fascio di elettroni contro uno di positoni. La curva mostra che la massa del mesone rho è di 774 ± 4 MeV e la sua ampiezza di risonanza di 111 ± 6 MeV. Maggiore è l'ampiezza di risonanza di queste particelle (note anche come risonanze) e minore la loro vita media. La maggior parte dei rho decadono prima di percorrere pochi raggi nucleari. Picchi simili, ma molto piú piccoli, corrispondono alle masse del mesone omega e del mesone phi.

ne. Questa relazione fra assorbimento e diffrazione è nota come « teorema ottico »; si applica l'aggettivo « iconale » ai modelli in cui questo teorema è importante. Il termine iconale sembra particolarmente appropriato nel fatto che i fotoni diffratti formano un'immagine del bersaglio mentre i fotoni assorbiti rendono conto dell'ombra. Nella sua essenza, quindi, il teorema, ottico collega l'immagine con l'ombra.

L'applicazione del teorema ottico ai fotoni di alta energia è molto bene illustrata da esperimenti recenti sulla diffusione elastica (diffrazione) di fotoni di alta energia su protoni. Qui « elastico » significa che non si producono altre particelle e che sia il fotone sia il protone sopravvivono all'esperimento. Al diminuire dell'angolo di diffusione, la probabilità della diffusione si avvicina alla probabilità prevista dal teorema ottico. Questa previsione è a sua volta basata sulle probabilità, misurate separatamente, dell'assorbimento di fotoni da parte dei protoni, eventi in cui né il fotone né il protone sopravvivono. Perciò la diffusione è collegata all'assorbimento: l'immagine è collegata all'ombra.

Benché il modello a dominanza vettoriale descriva un certo numero di processi che coinvolgono fotoni reali o virtuali e mesoni vettoriali, esso collega specificamente le quantità seguenti: primo, le costanti che indicano la forza dell'accoppiamento di un fotone con ogni mesone vettoriale; secondo, la probabilità che un fotone colpendo un nucleone (protone o neutrone) produca ciascuno dei mesoni vettoriali; terzo, la probabilità che il mesone vettoriale colpendo un nucleo interagisca con esso; quarto, la probabilità totale che un adrone assorba un fotone, in altre parole, la probabilità totale che un fotone colpendo un nucleo interagisca fortemente. Il modello viene verificato misurando queste probabilità e le costanti di accoppiamento in vari esperimenti.

Sono oggi noti tre mesoni vettoriali con stranezza zero: il rho, l'omega e il phi. La carica del rho può essere positiva, negativa o nulla; l'omega e il phi esistono solamente con carica nulla. Poiché i fotoni hanno carica nulla, possono accoppiarsi fortemente solamente con i mesoni vettoriali neutri. Le regolarità e le posizioni vuote osservate nelle classificazioni delle particelle elementari ispirano i tentativi per spiegarle e predirle. Uno di questi tentativi è il modello a quark, ideato indipendentemente da Murray Gell-Mann e da George Zweig dell'Istituto di Tecnologia della California. Essi avanzano l'ipotesi che le particelle elementari siano combinazioni di subunità, forse solamente matematiche, chiamate quark. Il modello a quark predice che il fotone si comporta come fosse per il 75 % un rho, per l'8 % un omega e per il 17 % un phi. Ciò significa che in un gran numero di interazioni il fotone si comporterà come un rho nel 75 % dei casi, come un omega nell'8 % dei casi e come un phi nel 17 %. Perciò il rho è il più importante dei mesoni vettoriali nel mediare le interazioni fotone-adrone; perciò si parla spesso di dominanza del rho invece che di dominanza dei mesoni vettoriali. I mesoni vettoriali possono essere facilmente prodotti sia dagli adroni sia dai fotoni, in modo che si possono confrontare i vari meccanismi di produzione. Le vite medie dei mesoni vettoriali, benché molto maggiori di quelle di un fotone virtuale, sono molto brevi. Nessun mesone vettoriale vive abbastanza a lungo da lasciare una traccia in una camera a bolle o in una camera a scintille, e ognuno deve essere osservato indirettamente attraverso le particelle a vita più lunga in cui decade. Il rho decade di solito in due pioni, l'omega in tre pioni e il phi in due kaoni.

La massa di un particolare mesone vettoriale dev'essere calcolata in base alle masse, energie e momenti dei suoi prodotti di decadimento. Sperimentalmente, anche quando gli errori nelle misure sono trascurabili, le masse calcolate da un campione dei decadimenti dello stesso mesone vettoriale non coincidono ma formano invece una distribuzione di risonanza, una curva a campana con un valore centrale della massa (*si veda la figura a pag. 70*). Ciò in base al principio di indeterminazione, che richiede solamente che il prodotto della vita media e dell'ampiezza della distribuzione risonante di una massa sia costante. Più ampia è la distribuzione di massa, più breve la vita media; più stretta la distribuzione di massa, più lunga la vita media.

Le prime previsioni sull'esistenza dei mesoni vettoriali furono fatte nel 1957 da Yoichiro Nambu, e da William R. Frazer e Jose R. Fulco due anni dopo, per spiegare la struttura diffusa osservata per protoni e neutroni, particolarmente negli esperimenti di diffusione elettrone-nucleone effettuati da Robert Hofstadter e i suoi colleghi alla Stanford University. Il nome « mesone vettoriale » deriva dal momento angolare intrinseco, o spin, di queste particelle, che è unitario e che si trasforma da un sistema di coordinate spaziali a un altro come un vettore. Un altro numero quantico, la parità intrinseca, è negativo. Lo spin e la parità, come tutti gli altri numeri quantici interni, sono identici per i fotoni e i mesoni vettoriali.

Il metodo migliore per studiare la composizione del fotone in mesoni vettoriali è quello di far collidere un fascio di elettroni con uno di positoni, il che isola la produzione dei mesoni vettoriali dalle altre interazioni. Esperimenti simili sono stati effettuati a Orsay in Francia e a Novosibirsk nell'URSS. L'annichilazione di una coppia elettrone-positone genera frequentemente un solo fotone virtuale che può immediatamente materializzarsi in uno qualsiasi dei mesoni vettoriali che lo compongono. Al solito, la creazione di un particolare mesone vettoriale è segnalata dalla sua massa che si può calcolare dalle energie e dai momenti dei suoi prodotti di decadimento che vengono rivelati dalle camere a scintille e dai contatori disposti attorno alla zona di interazione.

L'annichilazione di un positone e di un elettrone in un solo fotone virtuale è puramente elettromagnetica e può essere calcolata con buona precisione. Inoltre, tutti i mesoni vettoriali prodotti dai fotoni virtuali decadono immediatamente, per cui è solamente necessario misurare quale frazione di questi decadimenti viene rivelata e di calcolare l'efficienza dell'apparato di rilevamento per determinare la probabilità che si materializzi un fotone virtuale. La composizione percentuale dei fotoni trovata negli esperimenti di collisione dei fasci di Orsay è di 75 ± 2 rho, di 10 ± 2 omega e di 15 ± 2 phi, in buon accordo con le previsioni del modello a quark. Gli esperimenti misurano direttamente anche le masse centrali e le ampiezze delle distribuzioni risonanti di massa, che in altri esperimenti sembrano influenzate dalla presenza di adroni nel bersaglio. Gli esperimenti di collisione dei fasci effettuati a Orsay e Novosibirsk sono non solamente brillanti sul piano tecnico ma hanno il notevole vantaggio di non avere adroni nel bersaglio.

La produzione di mesoni vettoriali mediante fotoni reali è stata studiata in una grande varietà di esperimenti. Tutti i metodi sopradescritti per generare fasci di fotoni di alta energia sono stati usati con successo: radiazione di *bremsstrahlung*, annichilazione elettrone-positone e diffusione Compton inversa dei fotoni laser per mezzo di elettroni di alta energia. I rivelatori sono stati di tipo visualizzante (camere a bolle, camere a scintille e camere a scintillazione), oppure spettrometri magnetici di 2000 tonnellate che rivelano singole particelle di rinculo, come i protoni del bersaglio, da cui si può dedurre la produzione di risonanze.

Si può dire in generale che i dispositivi visualizzanti rendono possibile un'analisi dettagliata sia dei processi di produzione sia di quelli di decadimento, anche per gli eventi in cui sono implicate molte particelle finali. I dispositivi elettronici, quali gli spettrometri magnetici con vari tipi di contatori, danno informazioni meno particolareggiate su determinati eventi e non possono registrare bene gli eventi complessi che producono molte particelle. Essi hanno però un vantaggio importante, dato che possono essere impiegati con fasci intensi per accumulare un gran numero di eventi o per studiare processi che avvengono solo raramente e dopo che molti fotoni sono passati per il bersaglio.

Non è troppo difficile determinare quale dei tre mesoni vettoriali è stato prodotto analizzando i prodotti di decadimento. Si vorrebbe anche analizzare, però, ciò che succede quando un rho e un altro mesone vettoriale urtano in seguito contro un nucleone. Ma come si può sperare di studiare la diffusione rho-nucleone quando il percorso del rho durante la sua breve vita è troppo corto per essere visibile, tanto corto che il rho decade infatti normalmente in due pioni entro qualche raggio nucleare dal suo punto di produzione? La risposta, avanzata da Marc H. Ross e Leo Stodolsky e, indipendentemente, da Sidney D. Drell e James Trefil, è che si deve studiare la diffusione rho-nucleone all'*interno* stesso del nucleo in cui il rho è stato prodotto. La produzione del rho senza una susseguente diffusione rho-nucleone può essere studiata con il nucleo più semplice: il nucleo dell'idrogeno ordinario, che consiste di un singolo nucleone. I risultati ottenuti dall'idrogeno possono essere confrontati con quelli ottenuti da nuclei di diametro crescente, quali quelli del berillio, carbonio, alluminio, rame, argento e piombo, che mostrano in maniera sempre più marcata gli effetti della diffusione rho-nucleone e dell'assorbimento dei rho.

Negli esperimenti di fotoproduzione dei rho effettuati all'elettrosincrotrone di Amburgo (DESY), alla Cornell University e allo Stanford Linear Accelerator Center (SLAC) furono ottenuti dei dati con più di una dozzina di nuclei bersaglio diversi su una ampia gamma di energie fotoniche e angoli di produzione dei rho. In termini qualitativi, per piccoli angoli la fotoproduzione dei rho per nucleo del bersaglio aumenta con il numero dei nucleoni nel nucleo bersaglio fino a circa 64 nucleoni (rame) e rimane poi costante o anche diminuisce. L'aumento per nucleone indica che i nucleoni del bersaglio collaborano insieme coerentemente per produrre rho più efficacemente di quanto farebbe lo stesso numero di nucleoni indipendenti. Questo fenomeno è analogo all'interferenza costruttiva che si osserva con la luce di una sorgente coerente come il laser. Nei nuclei più pesanti un sempre maggior numero di rho viene assorbito attraverso la diffusione rho-nucleone prima che possa sfuggire. Questo assorbimento supera il vantaggio della produzione coerente di rho, e perciò la frequenza per nucleone del bersaglio non aumenta per elementi con più di 64 nucleoni (*si veda la figura a pag. 67*).

Il risultato sperimentale indica che la distanza caratteristica necessaria perché avvenga un'interazione rho-nucleone nella materia nucleare (cammino libero medio) è di circa tre fermi, cioè pari a circa il raggio di un nucleo di carbonio e quattro volte il raggio di un protone (un fermi è 10^{-13} cm). Quando i nuclei del bersaglio vengono bombardati con fasci di adroni convenzionali, quali protoni e pioni, invece che

Gli studi sull'assorbimento dei fotoni sono stati eseguiti con il dispositivo sperimentale qui mostrato, progettato da un gruppo dell'Università della California a Santa Barbara in collaborazione con lo SLAC. Positoni secondari da 19,5 GeV entrano da sinistra e colpiscono un bersaglio di tungsteno che funge da irradiatore di fotoni. A seconda dell'energia che i positoni perdono nell'irradiare i fotoni, essi sono deviati in modo diverso dal magnete di « marcatura ». Un magnete più forte, allineato col primo, devia i positoni che non hanno perso energia. I fotoni « marcati » (cioè quelli la cui energia è nota da quella perduta dai positoni che li hanno prodotti) viaggiano in linea retta finché non colpiscono il bersaglio predisposto. Se un fotone viene assorbito, esso genera dei pioni che sono registrati nei contatori di adroni; i fotoni che non sono assorbiti entrano nel contatore del fascio di fotoni. Questo esperimento serve molto bene a dimostrare il carattere adronico dei fotoni. I fotoni che producono coppie elettrone-positone o che sono diffusi dai neutroni non vengono rivelati come « eventi adronici »; essi passano attraverso i fori dei contatori.

Il comportamento adronico dei fotoni è dimostrato da questa raccolta di dati che mostra l'assorbimento totale dei fotoni da parte dei protoni negli eventi adronici. I risultati ottenuti per i fotoni virtuali negli esperimenti di diffusione anelastica dello SLAC (*punti in colore*) sono in buon accordo con un certo numero di altri esperimenti in cui sono stati usati fotoni reali. I picchi stretti sotto i 2 GeV e il continuo uniforme sopra quel valore duplicano esattamente la curva che si ottiene usando i pioni (adroni autentici). Perciò i dati ottenuti sia per i fotoni virtuali sia per quelli reali mostrano molto chiaramente che i fotoni possono avere un comportamento di tipo adronico.

con i « fasci di rho » generati entro i singoli nuclei, i risultati sono simili. Quando il mesone rho interagisce con gli adroni, perciò, esso si comporta proprio come un adrone.

Una previsione importante derivante dal modello a dominanza vettoriale riguarda la probabilità che un fotone di alta energia interagisca fortemente con un nucleone. In un simile processo tutta l'energia elettromagnetica del fotone si materializza in adroni. Anche in questo caso tutte le tecniche che si conoscono per produrre fotoni di energia grande e nota con precisione sono state usate in esperimenti di assorbimento totale.

Questi esperimenti, che si estendono fino a un'energia di 18 GeV, sono in buon accordo fra di loro (*si veda la figura in alto in questa pagina*). Essi portano alla conclusione generale che i protoni assorbono fotoni reali e virtuali allo stesso ritmo e che, benché i protoni assorbano gli adroni 200 volte piú prontamente di quanto non assorbano i fotoni, la variazione dell'assorbimento in funzione dell'energia è la stessa in entrambi i casi. In breve, siano i fotoni reali o virtuali, il loro ritmo di assorbimento da parte dei protoni è coerente con una loro natura adronica. Ciò tende a confermare l'importanza del processo a due fasi: un fotone si materializza in un mesone vettoriale, che a sua volta interagisce con un adrone.

La similarità fra protoni e neutroni è dimostrata dagli esperimenti Santa Barbara-SLAC che hanno misurato la capacità dei due nucleoni di assorbire i fotoni. La teoria prevede che se i fotoni agiscono come adroni, le due particelle dovrebbero assorbire fotoni di altissima energia allo stesso ritmo. La previsione è verificata se si estrapola la curva all'infinito. La piccola divergenza delle curve che appare per basse energie fotoniche può significare che all'assorbimento partecipano altri mesoni.

Il modello a dominanza vettoriale prevede che a energie molto alte l'assorbimento di fotoni da parte dei protoni e dei neutroni dovrebbe essere identico. Il confronto è stato fatto fino a 18 GeV presso lo SLAC da parte dei gruppi dell'Università della California a Santa Barbara e dello stesso SLAC. In questi esperimenti in collaborazione furono lanciati dei fotoni contro bersagli di deuterio, di idrogeno e di altri atomi. Mentre il nucleo dell'idrogeno è composto da un solo protone, il nucleo del deuterio è composto da un neutrone e da un protone. Ne risulta che l'assorbimento di un fotone da parte di un neutrone è quasi uguale alla differenza fra l'assorbimento da parte del deuterio e l'assorbimento da parte dell'idrogeno (bisogna fare una correzione del 5 % per l'ombra che un nucleone fa all'altro nel deuterio). Se si estrapolano questi risultati fino a un'energia infinita dei fotoni, gli assorbimenti da parte dei neutroni e dei protoni sono identici, come ci si attendeva. A energie minori l'assorbimento da parte dei protoni è circa 10 volte maggiore dell'assorbimento da parte dei neutroni, il che fa supporre che vi partecipino mesoni non vettoriali.

È nell'assorbimento totale dei fotoni di grande energia da parte di grandi nuclei, però, che il carattere adronico della luce ha le conseguenze piú stupefacenti. Qui, piú che in ogni altro fenomeno, la natura della luce dipende dal fatto che si osservino le sue interazioni adroniche, oppure quelle elettromagnetiche. Come abbiamo spiegato, il ritmo assoluto di assorbimento di fotoni da parte di protoni o neutroni è circa 200 volte minore di quello di assorbimento di rho o di pioni da parte dei protoni. Ciò non rappresenta una sorpresa, in quanto il primo passo – un fotone che si materializza in mesone vettoriale – del processo di interazione in due fasi è piuttosto improbabile. D'altra parte, il cammino libero medio dei fotoni che corrisponde a questo ritmo di assorbimento osservato è di circa 700 fermi, cioè 200 volte maggiore dei tre fermi previsti per i rho. È come se fossero necessari 700 strati di materia nucleare per assorbire la luce nella sua forma elettromagnetica, mentre tre strati dello stesso materiale sono sufficienti per fermare gli adroni. Perciò i fotoni passano facilmente attraverso grandi nuclei ma gli adroni penetrano appena la superficie, interagendo raramente con i nucleoni dell'interno.

Poiché i fotoni passano indisturbati attraverso tanto spazio nucleare, ci si attenderebbe che essi «illuminino» tutti i nucleoni di un dato nucleo. L'assorbimento dei fotoni sarebbe allora proporzionale al numero totale di nucleoni in ciascun nucleo e quindi al volume del nucleo. Al contrario, è noto che l'assorbimento degli adroni è proporzionale al numero dei nucleoni che si trovano sulla superficie ed è perciò proporzionale all'area della superficie nucleare, come se ogni nucleo fosse un disco nero. Paradossalmente, il modello a dominanza vettoriale e il teorema ottico portano alla previsione che il rapporto fra l'assorbimento nucleare dei fotoni e quello degli adroni dipenda solo dalla probabilità che un fotone si materializzi come un adrone e che sia identico per ogni nucleo, con ambedue i processi variabili in funzione dell'area della superficie e non del volume nucleare. Perciò occorre pensare che, benché i nuclei siano quasi trasparenti ai fotoni di grande energia, i nucleoni della superficie possano oscurare completamente quelli all'interno.

Questo è il tipo di paradosso che la natura fornisce continuamente a coloro che hanno troppa confidenza del loro senso comune. Per esempio, nel 1819 il matematico francese Siméon Denis Poisson demolí la teoria ondulatoria della luce mediante una *reductio ad absurdum*. Egli mostrò in maniera matematicamente impeccabile che la teoria ondulatoria di Fresnel prevede che l'ombra di un disco opaco illuminato da una sorgente puntiforme abbia una macchia luminosa nel suo centro, il che sembra chiaramente assurdo. Ci spiace per monsieur Poisson ma l'ombra causata da un disco simile, per esempio da una monetina illuminata da una sorgente puntiforme, ha effettivamente una zona centrale luminosa.

Per capire il paradosso della superficie osservato con i fotoni che interagiscono fortemente, dobbiamo invocare ancora il principio di indeterminazione. Piú esattamente, la posizione sulla traiettoria del fotone nella quale il fotone è convertito in un mesone vettoriale è indeterminata di una quantità proporzionale all'energia del fotone. Quando l'indeterminazione è grande in confronto con il cammino libero medio degli adroni entro i nuclei (come accade per i fotoni di grande energia), il fotone deve essere stato convertito in un adrone ben prima di raggiungere la superficie nucleare, nel qual caso si può avere assorbimento ed effetto d'ombra. Il ragionamento si può invertire. Supponiamo che non vi sia l'effetto ombra; allora la conversione in un rho di un fotone deve essere avvenuta immedia-

La schermatura dei nucleoni da parte di altri nucleoni accade quando nuclei atomici di varie dimensioni vengono bombardati con fotoni. Se i fotoni si comportassero in maniera puramente elettromagnetica, l'effetto ombra sarebbe nullo. La banda colorata indica la percentuale di schermatura per nucleone medio prevista dal modello della « dominanza vettoriale » del comportamento del fotone. Gli esperimenti del gruppo di Santa Barbara allo SLAC danno valori alquanto piú bassi (*punti*) per il deuterio, il carbonio, il rame e il piombo. I valori sono delle medie per fotoni fra 7 e 18 GeV.

tamente di fronte al punto di assorbimento. Ciò distruggerebbe l'indeterminazione nel punto di conversione; perciò dev'esservi l'effetto ombra!

Significa ciò che un grosso nucleo assorbe la maggior parte dei fotoni che lo colpiscono? Tutt'altro: quasi tutti i fotoni lo attraversano indisturbati. La probabilità che un fotone mostri un comportamento adronico e interagisca fortemente è, però, limitata in effetti ai nucleoni superficiali e non si estende ai nucleoni dell'interno. Il fotone, dopo essere penetrato nella superficie nucleare, è in pratica un fotone « nudo » spogliato del suo carattere adronico. Supponiamo che un fotone passi successivamente attraverso due nuclei distanti: il primo fa ombra al secondo? No. Alle energie raggiungibili oggi la distanza, stabilita dal principio di indeterminazione, nella quale il fotone rimane nudo è molto minore della distanza fra i nuclei. Perciò la schermatura di un nucleo atomico da parte di un altro avviene raramente.

La previsione dell'effetto ombra da parte dei nucleoni di superficie segue direttamente dal ragionamento sul principio di indeterminazione e dal carattere adronico della luce, indipendentemente da ogni particolare sul modello a dominanza vettoriale. Una prova sufficiente del carattere adronico è data perciò dagli studi di assorbimento totale dei fotoni su nuclei di dimensioni crescenti. I fotoni possono essere considerati puramente elettromagnetici solo se il nucleo rimane trasparente, solo se tutti i nucleoni sono illuminati e solo se l'assorbimento è proporzionale al numero totale di nucleoni.

I risultati ottenuti nell'esperimento dei gruppi di Santa Barbara e dello SLAC con bersagli di idrogeno, deuterio, carbonio, rame e piombo mostrano che l'assorbimento di fotoni non è proporzionale né al numero totale di nucleoni né a quello dei nucleoni di superficie; i valori cadono approssimativamente circa all'interno (*si veda la figura a pag. 75*). Un certo grado di effetto ombra è certamente presente, per cui è confermato il carattere adronico della luce. Da un punto di vista qualitativo, perciò, i risultati sono consistenti con il modello a dominanza vettoriale. Dal punto di vista quantitativo, l'entità della schermatura è minore di quella che la dominanza vettoriale, assieme ai risultati di altri esperimenti, farebbero pensare. Significa ciò che vi sono altri mesoni vettoriali che si accoppiano al fotone oltre al rho, all'omega e al phi? Forse i fotoni diventano mesoni vettoriali meno facilmente di quanto si pensi, o forse i mesoni vettoriali vengono assorbiti meno prontamente nella materia nucleare di quanto non facciano supporre gli esperimenti odierni. La risposta a questi problemi, come pure il destino ultimo del modello a dominanza vettoriale, dipendono forse dagli esperimenti già progettati per l'acceleratore da 76 GeV di Serpukhov, vicino a Mosca, e per l'acceleratore da 500 GeV in fase di avanzata costruzione a Batavia negli USA.

La diffusione elastica dei fotoni su protoni varia con l'angolo di diffusione. La curva collega i punti dei dati ottenuti allo SLAC per fotoni di 16 e 17 GeV. Al diminuire dell'angolo di diffusione la probabilità della diffusione si avvicina a quella del « teorema ottico ». All'avvicinarsi a zero dell'angolo di diffusione la maggior parte dei fotoni incidenti viene sempre più diffusa, finché alla fine la diffusione a cui un fotone sopravvive è collegata dal teorema ottico alla probabilità del suo assorbimento.

L'ecosistema del parco di Serengeti

Grandi branchi di zebre, gnu e gazzelle di Thomson migrano attraverso le pianure della Tanzania. Queste migrazioni sono strettamente legate alla presenza nei pascoli della Tanzania di particolari specie vegetali

di Richard H.V. Bell

Le migrazioni stagionali degli erbivori attraverso la savana africana offrono uno spettacolo naturale di grande interesse. È chiaro che queste migrazioni, a cui spesso partecipano parecchie migliaia di animali, sono legate all'avvicendarsi delle stagioni secche a quelle delle piogge e viceversa. Un'altra influenza, peraltro non molto chiara, è emersa dai lavori condotti nell'ultima decade nei vasti pascoli della regione di Serengeti, nella parte settentrionale della Tanzania. Accade infatti che la vegetazione e gli animali si influenzano a vicenda, cosicché si verifica una successione caratteristica di specie erbivore in stretta connessione con gli avvicendamenti stagionali.

Studiando questo fenomeno nel parco di Serengeti, abbiamo notato che le principali specie migratrici (gli gnu, le zebre e le gazzelle di Thomson) hanno particolari ragioni fisiologiche quando scelgono per nutrirsi alcuni tipi di tessuti vegetali e di erbe (i principali tessuti sono le foglie, le guaine e gli steli) e cosí hanno la tendenza a migrare quando quel particolare tipo di vegetazione incomincia a scarseggiare nella zona in cui si trovano. È quindi la maggiore o minore disponibilità di una particolare specie vegetale che condiziona i loro spostamenti.

Nel Serengeti le prime a migrare sono le zebre seguite poco dopo rispettivamente dagli gnu e dalle gazzelle di Thomson. Anche un'antilope (*Damaliscus corrigum*), una specie semimigratoria assai diffusa e abbondante nella zona, tende ad associarsi alle zebre.

Il parco nazionale di Serengeti copre un'area di 14 500 km² di praterie e boscaglie isolate e forma la parte centrale del complesso ecologico di Serengeti di 23 300 km². Sia il parco nazionale, sia le zone di caccia controllata che fanno parte del complesso ecologico, sono sotto il diretto controllo del governo della Tanzania. In quest'area sono state condotte già da circa 14 anni delle ricerche ecologiche, a partire da un rilievo generale eseguito nel 1957 da W. H. Pearsall dell'University College di Londra, e comprendente uno studio dello zoologo tedesco Bernhard Grzimek e di suo figlio Michael, eseguito mediante ricognizioni aeree.

Nel complesso ecologico di Serengeti vivono piú di mezzo milione di animali selvatici, appartenenti quasi tutti a specie migratrici; il complesso ecologico è stato definito da R.M. Watson, del Progetto di ricerca Serengeti, come l'area occupata regolarmente dalle popolazioni migratrici nel corso degli spostamenti stagionali. L'aspetto piú importante nell'ecosistema di Serengeti è l'utilizzazione del pascolo da parte degli erbivori ungulati, che costituiscono piú del 90 per cento della biomassa dei mammiferi. In ordine di grandezza, le specie piú numerose di ungulati sono il bufalo (*Syncerus caffer*), la zebra (*Equus burchelli*), lo gnu (*Connochaetes taurinus*), il damalisco (*Damaliscus corrigum*) e la gazzella di Thomson (*Gazella thomsoni*).

Per capire come i vari animali utilizzino lo strato erboso si deve innanzitutto prendere in esame la struttura cellulare vegetale.

Ogni cellula vegetale è circondata da una parete abbastanza resistente, costituita principalmente da cellulosa; all'interno vi è il citoplasma, che è formato principalmente di proteine e di carboidrati solubili. La struttura di qualsiasi parte di una pianta, sia essa un tessuto vegetale, un organo o tutta la pianta, dipende dalla consistenza delle pareti cellulari. Nei tessuti giovani in via di sviluppo, che debbono accrescersi e che non debbono sostenere peso, le pareti cellulari sono sottili e abbastanza elastiche: esempi di questi tessuti si trovano nelle foglie e nei frutti delle piante erbacee e di altre piante da fiore. Nei tessuti piú vecchi, specialmente quelli che hanno il compito di sostenere un peso, le pareti cellulari si ispessiscono e s'induriscono per la presenza di altre sostanze, in particolare la lignina, mentre il citoplasma si riduce: questo è il caso d'una pianta erbacea ad alto fusto o d'un cespuglio. Alla fine la parete cellulare lignificata, accrescendosi verso l'interno, riduce la cavità interna e il tessuto che si forma è il legno.

Per gli erbivori, la suddivisione del loro alimento vegetale in parete cellulare, ossia cellulosa e citoplasma, è importante perché influenza nettamente le caratteristiche principali del cibo: il contenuto proteico e energetico.

I mammiferi erbivori non posseggono enzimi digestivi propri in grado di attaccare le membrane cellulosiche delle piante di cui si nutrono. Questi animali suppliscono poi a tale deficienza utilizzando a questo scopo alcuni microrganismi (batteri e protozoi) che vivono nel loro intestino e nel loro stomaco e che producono enzimi capaci di digerire la cellulosa. Anche con l'aiuto di questi microrganismi, tuttavia, la demolizione della cellulosa è lenta; quando poi è presente la lignina la cosa è ancora piú ardua. Perciò la parete cellulare costituisce un ostacolo fisico all'estrazione del contenuto cellulare, specialmente in quei tessuti in cui la parete cellulare è spessa e lignificata.

Non appena la cellulosa è stata idrolizzata, diviene una fonte d'energia per l'animale. Altra energia può essere ricavata dal citoplasma. La quantità totale di alimento energeticamente utile ricavato dalla pianta è superiore alla quantità fornita dalle proteine, che provengono solo da alcune parti cellulari e sono particolarmente scarse nei tessuti in cui le cellule hanno pareti spesse.

Per queste ragioni gli erbivori sono obbligati a ricavare le proteine da un alimento che contiene un eccesso di carboidrati: questa situazione crea un problema che viene superato in due fasi: quando il cibo viene inghiottito e durante la digestione; i due momenti sono diversi a seconda che l'erbivoro sia un ruminante (bovini, ovini, antilopi, cervidi) o un non ruminante (cavalli, zebre, elefanti). Schematizzando, si può dire che l'apparato digerente dei ruminanti sfrutta al massimo le proteine, a spese di un eccesso di energia.

Il ruminante giunge a questo risultato mediante una fermentazione pregastrica (il luogo dove la cellulosa fermenta è situato piú in alto del luogo dove le proteine vengono assorbite) e con un ciclo di rigurgito e masticazione. Il meccanismo funziona in questo modo: il cibo ingerito all'inizio è masticato e sottoposto agli enzimi salivari: quando è inghiottito per la prima volta, i suoi componenti sono ancora piuttosto grossi e molte cellule vegetali sono ancora intatte. Il rumine funziona come un filtro, separando i pezzi secondo il loro peso specifico: infatti le particelle di maggiori dimensioni sono anche quelle di piú basso peso specifico, perciò galleggiano sopra il liquido del rumine e vengono trasferite nel reticolo, e da qui vengono rigurgitate. Il ciclo si ripete finché le cellule vegetali si sono sminuzzate e decomposte completamente: allora affondano nel liquido del rumine, da cui passano poi al vero e proprio stomaco e quindi all'intestino.

Tuttavia la velocità di passaggio del cibo attraverso l'intestino è limitata dalla velocità di demolizione del materiale nel rumine, che è a sua volta dipendente in larga misura dalla quantità e dalla qualità delle pareti cellulari nel cibo. Quando l'alimento contiene una grande quantità di pareti cellulari integre, specialmente se lignificate, la velocità di passaggio è lentissima: perciò anche se le proteine possono essere estratte, la loro assimilazione risulta rallentata.

Quando i costituenti della parete cellulare raggiungono una certa concentrazione nel cibo, il ruminante non riesce piú ad assimilare proteine in quantità sufficiente al suo mantenimento. In questa situazione l'animale è costretto a scegliere vegetali con pareti cellulari sottili e a concentrazione proteica elevata: foglie, frutti e germogli delle erbe e di altre piante dello strato erboso. Tuttavia, nel medesimo tempo, il meccanismo di fermentazione pregastrica rimette in circolo e ricostituisce le sostanze alimentari, per assicurare la massima utilizzazione delle proteine estratte. I criteri alimentari dei ruminanti si basano quindi sull'efficacia dell'estrazione e dell'utilizzazione delle proteine.

Non si conoscono altrettanto bene invece i processi nutritivi degli erbivori non ruminanti. Gli studi piú approfonditi sono stati eseguiti sul cavallo domestico, animale molto affine alla zebra. A differenza dei ruminanti, il cavallo fermenta la cellulosa nell'intestino che ha uno sviluppo particolarmente lungo ed è ingrossato nella regione del colon; l'estrazione e l'assimilazione delle proteine avvengono invece nello stomaco, relativamente piccolo e semplice. L'animale quindi estrae le proteine senza la macerazione meccanica e l'attacco chimico che nei ruminanti demoliscono le pareti cellulari. L'estrazione delle proteine viene ancor piú diminuita, rispetto ai ruminanti, quando la parete cellulare è particolarmente ispessita. Inoltre le proteine estratte nello stomaco del cavallo vengono subito assimilate sotto forma di amminoacidi, e non sono soggette a quel complesso meccanismo di riciclaggio e ricostituzione che nel ruminante aumenta il rendimento dell'utilizzazione proteica.

Per tutte queste ragioni il cavallo ha, rispetto a un ruminante, minori possibilità di utilizzare le proteine contenute nel cibo. Quando il cibo è limitato, i ruminanti hanno maggiori probabilità di sopravvivenza rispetto agli equini. Allora, come riescono gli equini a mantenere la propria posizione nell'ecosistema? La risposta sta nel loro diverso apparato digerente: in questo apparato non vi è un meccanismo che imponga un limite alla velocità di passaggio del materiale attraverso l'intestino. Confrontando un non ruminante come il cavallo, con la mucca, che è ruminante, si constata che il materiale alimentare passa attraverso l'intestino del cavallo a una velocità doppia rispetto a quella dell'intestino della mucca (da 30 a 45 ore nel cavallo, da 70 a 100 ore nella mucca).

Per comprendere l'importanza di queste considerazioni basterà un semplice calcolo: se per estrarre proteine da un cibo il cavallo ha un rendimento pari ai due terzi rispetto a un rumi-

La regione di Serengeti è al confine tra la Tanzania e il Kenia, nell'Africa orientale. Comprende il parco nazionale di Serengeti, che ha una estensione di 14 500 km² di praterie ed è racchiuso in un complesso ecologico che ha un'area di 23 300 km².

BUFALO

447 CHILOGRAMMI

42 000 ANIMALI

ZEBRA

219

220 000

GNU

163

330 000

DAMALISCO

108

20 000

GAZZELLA DI THOMSON

16

150 000

Le principali specie di erbivori della regione di Serengeti sono raffigurate in questo schema in ordine di peso, espresso in kg e per una femmina adulta (*colonne verticali*). L'entità numerica è espressa a destra: ogni quadratino rappresenta 10 000 animali.

nante, ma sottopone a trattamento una quantità doppia di cibo in un certo periodo di tempo, l'assimilazione di proteine per unità di tempo rappresenterà i quattro terzi di quella di un ruminante. Perciò il cavallo (o la zebra) può mantenersi in vita con un'alimentazione che sarebbe troppo povera di proteine per un ruminante. In altre parole, l'animale non ruminante tollera più del ruminante la presenza di pareti cellulari nell'alimento, ma deve introdurre una grande quantità di materiale per potersi sfamare con un cibo di questo genere: perciò la sua alimentazione è molto meno selettiva di quella del ruminante.

Nonostante questa diversità riguardo al tipo di alimentazione, in realtà tra questi due gruppi di animali vi è qualche punto in comune, dovuto all'effetto delle dimensioni corporee sul metabolismo basale, e perciò sul fabbisogno di proteine e di alimenti energetici: piú piccolo è l'animale, piú alto è in genere il metabolismo basale. Esaminando gli aspetti ecologici occorre tener presente sia il fabbisogno alimentare relativo (per unità di peso e per unità di tempo), sia quello assoluto (per animale e per unità di tempo).

Per il fabbisogno relativo, il metabolismo particolarmente vivace è causa negli animali piccoli di una maggiore necessità di proteine e d'energia per unità di peso e per giorno, rispetto ai grossi animali: il fabbisogno proteico giornaliero di una gazzella di Thomson, per unità di peso è, per esempio, da due a quattro volte maggiore di quello dello gnu. In conseguenza gli animali piú piccoli perdono peso piú rapidamente rispetto a quelli grandi, quando vengono sottoposti a un'alimentazione insufficiente: se quindi due specie di dimensioni differenti hanno il medesimo rifornimento di cibo, la specie di dimensioni maggiori soppianterà quella di dimensioni minori.

Benché il fabbisogno alimentare relativo sia maggiore in un animale di piccole dimensioni rispetto a quello di un animale grosso, la situazione s'inverte quando si considera il fabbisogno assoluto. Confrontando nuovamente una gazzella di Thomson e uno gnu, la prima ha un fabbisogno minimo pari al 20 per cento di quello del secondo. (Il calcolo presuppone qualità uguali di cibo; il fabbisogno della gazzella sarebbe ancora minore se essa fosse piú selettiva in fatto di cibo, rispetto allo gnu.) In altre parole, la gazzella ha a sua disposizione un tempo cinque volte maggiore di quello dello gnu per trovare e mangiare la quantità di cibo necessaria alla sua sopravvivenza. Le specie piú piccole si possono per-

LAGO VITTORIA

GOLFO DI SPEKE

TANZANIA

KENIA

NAIROBI

BANAGI

SERONERA

MONTE KILIMANGIARO

Il parco nazionale di Serengeti (*in colore*) è a est del Golfo di Speke sul Lago Vittoria. Il quartier generale del parco si trova a Seronera; fino al 1967, si trovava a Banagi. Le grosse linee grige rappresentano le strade principali esistenti nella regione.

SERENGETI SETTENTRIONALE

SERENGETI OCCIDENTALE

PIANORO DI SERENGETI

Le tre zone ecologiche sono caratterizzate da piovosità elevata ed erba alta (nord), da piovosità bassa ed erba bassa (pianoro di Serengeti), da piovosità media e erba di media altezza (ovest).

Aree occupate dagli animali migratori. Gli animali trascorrono la stagione delle piogge (da marzo a maggio) nel pianoro di Serengeti, la prima parte di quella asciutta a ovest e la fine a nord.

ciò permettere il lusso di essere piú esigenti di quelle di dimensioni maggiori: possono sostentarsi con riserve di cibo scarse e decisamente insufficienti per un animale piú grosso.

Un animale di piccole dimensioni sopravvive meglio di uno piú grosso, se si verificano queste due condizioni: forte assimilazione delle proteine e degli alimenti energetici per unità di peso corporeo, e bassa assunzione di proteine e di alimenti energetici per animale. A questi requisiti corrispondono bene i tessuti vegetali che contengono poca cellulosa: tali tessuti però hanno spesso una distribuzione sparsa e sono difficili da raggiungere.

Piú l'animale è di notevoli dimensioni, piú è facile la sua vita in condizioni opposte alle precedenti: l'animale infatti accetta bene i tessuti vegetali che contengono alte percentuali di cellulosa. Piú sono frequenti e accessibili questi tessuti, piú è grosso l'animale che può sostentarsi con essi. Questo è in particolar modo il caso degli erbivori non ruminanti, come la zebra, che sono costretti ad assumere molto cibo, ma che tollerano materiale con molta cellulosa.

Per poter vedere come le caratteristiche alimentari dei principali esponenti della fauna erbivora di Serengeti impongano loro un determinato uso delle riserve di cibo e spostamenti stagionali prefissati, è necessario esaminare a fondo la struttura e la distribuzione di queste riserve di cibo. Queste considerazioni si basano soprattutto sulla forma di crescita di ciascuna pianta erbacea. All'inizio della stagione umida il germoglio spunta da terra: in questo periodo iniziale la crescita si concentra in un punto che sta appena al di sotto dell'apice caulinare; si forma via via una serie di foglie che crescono attorno all'apice, racchiudendolo con le loro guaine: l'apice e il punto di accrescimento si trovano vicino al terreno. In questo stadio la pianta è formata da tessuti con pareti cellulari sottili e ha le foglie facilmente accessibili agli erbivori: l'animale inoltre può brucare le foglie senza distruggere l'apice caulinare, ben protetto nel suo punto di accrescimento.

La fase di crescita successiva si verifica più tardi, nella stagione umida: il fusto si estende al di sotto del punto di accrescimento, sollevando le foglie dal terreno, in risposta alla competizione per la luce che si svolge tra tutte le piante; questa crescita implica una grande riduzione del contenuto proteico nella pianta, a causa della struttura del fusto.

Nell'ultima fase il fusto, al di sopra della zona di accrescimento, si allunga per formare l'infiorescenza. Questa parte della pianta è costituita da tessuti simili al resto del fusto, nel senso che ha un rapporto elevato tra costituenti della parete cellulare e contenuto cellulare. A questa regola fanno eccezione i semi prodotti dall'infiorescenza, ma gli erbivori in genere li trascurano.

Quindi, alla fine della stagione delle piogge, lo strato erboso presenta una serie di livelli che hanno un diverso valore alimentare e una diversa accessibilità da parte degli animali. Il livello superiore è costituito dai culmi delle graminacee, portatori di infiorescenze, scarsi in contenuto proteico; il livello medio è formato dagli steli e dalle foglie delle erbe piú alte; piú sotto si trovano le foglie delle erbe piú umili e i giovani germogli: le vere e proprie erbe, come il trifoglio in un prato da falciare, stanno nello strato inferiore. La conseguenza piú importante di questa stratificazione è che la piú alta concentrazione di proteine si trova nel livello piú basso, quasi inaccessibile agli erbivori a causa del fitto strato di fusti erbacei che lo sovrasta.

Questa eterogeneità di alimenti è importante per gli erbivori e comporta differenze di composizione chimica e di accessibilità delle riserve di cibo. Abbiamo esaminato il contenuto dello stomaco di quattro o cinque tra le principali specie erbivore di Serengeti (escludendo il bufalo) per analizzare la possibilità che gli erbivori compiano una scelta tra i differenti livelli dello strato erboso.

Secondo il nostro procedimento, abbiamo suddiviso il contenuto gastrico a seconda che la pianta fosse una monocotiledone (per esempio una graminacea) o una dicotiledone (in genere le altre piante dello strato erboso) e anche a seconda della parte: frutto, fusto, foglia, guaina. Delle quattro specie esaminate, solo lo stomaco di una, la gazzella di Thomson, conteneva un'alta percentuale di pezzi di dicotiledoni: 40 per cento, principalmente frutti. In questo contenuto gastrico è presente un'alta percentuale di proteine e di alimenti energetici, come appunto ci si dovrebbe aspettare per questo piccolo animale. Il contenuto degli stomaci delle altre tre specie era costituito quasi interamente di graminacee e mostrava rapporti diversi tra foglie e fusti: lo stomaco dello gnu conteneva per lo piú foglie, quello della zebra un minimo quantitativo di foglie e quello del damalisco una quantità intermedia.

È probabile quindi che questi quattro animali scelgano il loro alimento a un livello differente dello strato erboso: le zebre si nutrono con il materiale contenente la piú alta percentuale di cellulosa, nella parte piú alta dello strato erboso; il damalisco e lo gnu scelgono un alimento contenente piú proteine, nello strato di foglie sottostante; le gazzelle, infine, cercano a livello del suolo i frutti ad alto contenuto proteico. Tranne che per il damalisco, questa serie corrisponde a un ordine di grandezza decrescente.

Queste differenze di scelta, imposte dalle dimensioni e dalla fisiologia degli

In questi diagrammi sono registrate le variazioni delle popolazioni di erbivori migratori, in rapporto alle precipitazioni e all'altezza dell'erba, per un periodo di tre anni. Le cifre sono state ottenute nella zona ovest del Serengeti. I massimi delle curve indicano il passaggio delle principali specie, all'inizio della stagione asciutta,

ALTEZZA DELL'ERBA

BASSA

MEDIA

ALTA

ALBERI

BOSCHETTO D'ACACIE

PIANO DI STRARIPAMENTO

ARGINE ALLUVIALE

LINEA DI DRENAGGIO

La caratteristica topografica della regione di Serengeti è costituita da una serie di ondulazioni. La distribuzione dell'erba alta e bassa è influenzata dal drenaggio: le acque scorrono dalle alture e si depositano nelle depressioni, perciò l'erba in queste zone tende a essere piú alta. Dove l'erba è piú alta gli erbivori hanno difficoltà nel raggiungere le parti piú nutrienti della pianta.

animali, hanno un effetto importante sul comportamento degli ungulati nella prateria. Questo fatto fu scoperto sperimentalmente da G.W. Arnold dell'Università di Londra, con ricerche sulle pecore: secondo queste ricerche l'assunzione di cibo è nettamente limitata dalla presenza di piante indesiderate, che in questo caso erano culmi di graminacee. È chiaro che l'interferenza dei fusti è maggiore quando la pianta o la parte di pianta necessaria è situata in basso nello strato erboso: perciò è piú importante per le gazzelle che per le zebre. In generale l'interferenza è maggiore per tutte le specie quando lo strato erboso è formato da piante erbacee alte, costituite principalmente da fusti.

Nel parco di Serengeti e in gran parte della savana africana si può osservare una distribuzione tipica delle piante erbacee alte e di quelle basse; nessuna zona è perfettamente piatta: il terreno presenta sempre ondulazioni che variano in altezza da pochi centimetri a qualche centinaio di metri, mentre la distanza tra le quote piú alte varia da pochi metri a parecchi chilometri. La pioggia che cade sulle ondulazioni scende dai pendii nelle depressioni, cosicché le zone piú basse hanno un rifornimento d'acqua per periodi di tempo piú lunghi rispetto alle zone elevate e alle cime; le aree piú basse si ricoprono anche di terreno piú pesante, argilloso: per queste ragioni l'erba è piú alta nelle depressioni che sui pendii o sulle cime. La sequenza dei tipi di suolo e di vegetazione viene chiamata, in un ambiente di questo tipo, « sequenza a catena » (*si veda l'illustrazione in alto*).

Questa disposizione schematica delle riserve alimentari ha un effetto predominante sugli spostamenti stagionali degli erbivori; nella stagione delle piogge, quando le graminacee sono alla loro massima altezza, tutti gli erbivori si concentrano in gruppi misti nelle zone sopraelevate dove l'erba è meno alta; brucando e calpestando nel medesimo tempo l'erba, gli animali riducono le graminacee a uno stadio giovanile, col fusticino breve in via di sviluppo; in queste condizioni essi non hanno nessuna difficoltà a mangiare le foglie e perciò si procurano una quantità di proteine sufficiente per il loro mantenimento. Si può dire anzi che in questo periodo vi è un sovrappiú di cibo: è il periodo in cui le femmine partoriscono e allattano i piccoli. La parte piú alta delle ondulazioni del terreno presenta inoltre il vantaggio di avere un buon drenaggio e di essere meno cedevole sotto gli zoccoli rispetto alle depressioni, spesso impregnate d'acqua.

Nella stagione asciutta l'erba cessa di crescere, e quella corta delle zone sopraelevate sparisce gradualmente: poiché la quantità di cibo a disposizione in queste aree diminuisce, le diverse specie di erbivori si trovano nell'impossibilità di provvedere al proprio mantenimento e sono costrette a discendere a turno nell'erba alta delle depressioni. I primi a muoversi sono gli animali piú grossi (bufalo e zebra), poi anche gli animali piú piccoli (damalisco, gnu e gazzella di Thomson) si spostano. La discesa ordinata delle specie nelle depressioni del terreno si può definire successione di pascolo; l'ordine della successione corrisponde all'ordine di selettività nell'alimentazione e, prescindendo dal damalisco, all'ordine di peso corporeo.

Nelle erbe piú alte delle depressioni è piú difficile per gli erbivori scegliere le parti con un sufficiente contenuto proteico a causa della preponderanza dei fusti: questo si verifica soprattutto per gli animali piú piccoli, che tendono a scegliere parti del livello inferiore dello strato erboso. L'attività dei primi membri della successione nel demolire il fitto strato di culmi, brucando e calpestando il terreno, è perciò utilissima per gli ultimi membri: l'interdipendenza è dimostrata dall'associazione delle specie erbivore in gruppi misti e può essere espressa quantitativamente con un diagramma di coefficienti di associazione, che raggiungono i massimi nella stagione delle piogge e verso la fine della stagione asciutta.

La conclusione ricavata è che i primi membri della successione preparano la vegetazione per i membri successivi: una riduzione numerica di una specie, a causa di malattia o di caccia, può cagionare una riduzione numerica di una altra specie. Per esempio, si può prevedere che dopo una diminuzione nella popolazione delle zebre si verifichi una diminuzione nella popolazione degli gnu piú piccoli.

Finora il mio discorso si è limitato agli spostamenti stagionali localizzati nel nord-ovest del complesso ecologico di Serengeti: altri ricercatori hanno riferito su situazioni simili in altre regioni dell'Africa orientale. Ma lo spettacolo piú interessante di Serengeti è la migrazione massiccia di zebre, gnu e gazzelle di Thomson in una zona vasta circa come la Toscana.

Queste migrazioni sono in realtà equivalenti ai movimenti localizzati, ma in scala molto maggiore. Per una serie di fattori geologici e climatici la regione di Serengeti si può suddividere in tre zone, sistemate a triangolo, che rappresentano in scala maggiore le ondulazioni topografiche della savana. A sud-est l'enorme distesa di erba bassa

del pianoro di Serengeti, caratterizzato da scarse precipitazioni e da un terreno formato da ceneri vulcaniche, corrisponde alle zone sopraelevate, coperte di erba bassa, dove gli animali si raggruppano nella stagione delle piogge.

Nella parte occidentale le precipitazioni sono superiori e l'erba è piú alta: in questa zona convergono i migratori all'inizio della stagione asciutta. Infine, la zona settentrionale è quella dove le precipitazioni sono massime e dove le erbe hanno l'altezza maggiore: i migratori si spostano verso questa località alla fine della stagione asciutta. Come negli spostamenti stagionali localizzati, le migrazioni sono caratterizzate da una successione di pascolo: prima vengono le zebre, seguono gli gnu e infine appaiono le gazzelle di Thomson; questa successione si ripete (con variazioni) in tutta la savana africana.

Manca solo nel quadro generale un elemento importante: l'uomo. Esso ha fatto parte di questo sistema già dai suoi primi giorni come primate predatore, vicino ai grandi branchi di selvaggina, nonostante che le sue armi fossero ancora inadeguate; ha trascorso un periodo in cui era cacciatore, e si serviva di arco, frecce e lance; è stato pastore e si è basato, per il proprio sostentamento, sugli animali domestici che condividevano la prateria con gli animali selvatici; è stato poi agricoltore, e da allora è stato in competizioni con gli animali selvatici. Infine è diventato viaggiatore e turista, ed è evaso dalle città per rivisitare la savana, sua patria originaria.

L'influenza dell'uomo pastore sulla prateria, allevando e pascolando il bestiame domestico, è antica. L'uomo ha sempre voluto mantenere bassa l'erba dei suoi pascoli. In altre parole, il pastore è stato uno dei membri della successione di pascolo, non in competizione con gli animali selvatici, ma in collaborazione con essi. Infatti si può dimostrare che oggi, nella regione di Serengeti, i maggiori concentramenti di animali selvatici si hanno nelle zone di attività pastorale, presente o passata. Per contro, la coltivazione e la recinzione dei campi ha un'influenza negativa sulle migrazioni degli animali: questa fase dell'attività umana sta infatti provocando una riduzione del numero e dell'area di distribuzione della fauna della savana. Ora l'uomo in quanto scienziato, cioè nell'ultima fase dell'evoluzione umana, deve comprendere in senso globale il funzionamento dell'ecosistema e quindi suggerire schemi d'intervento, tenendo presente però sempre le diverse necessità del pastore, dell'agricoltore e del turista.

Vie nervose del cervello

Nuove tecniche, combinando vari metodi di colorazione con l'uso del microscopio ottico ed elettronico, rendono possibile la ricerca dettagliata delle connessioni tra le cellule e dei circuiti di conduzione del cervello

di Lennart Heimer

Il funzionamento del cervello è basato su un sistema di miliardi di cellule nervose fisicamente organizzate, tanto per chiarire, secondo uno schema di collegamenti elettrici. Le cellule e i vari centri funzionali del cervello sono collegati da un elaborato circuito di vie e interconnessioni formanti una rete di comunicazione. Per poter comprendere nei minimi particolari come il cervello lavori, bisogna avere una chiara cognizione di questo schema di collegamenti elettrici.

Naturalmente questo schema da solo non può spiegare le attività della mente umana, ma un quadro esplicativo del sistema di collegamenti è una condizione necessaria per tale conoscenza e per piú di un secolo gli studiosi hanno cercato di tracciare una mappa delle vie nervose attraverso lo studio di sezioni di tessuto. In questi ultimi anni l'introduzione di nuove tecniche, secondo le quali viene combinato l'uso del microscopio elettronico con quello ottico, ha reso possibile l'indagine dettagliata dei collegamenti del cervello.

La comunicazione nel sistema nervoso consiste nella trasmissione di segnali elettrici da una cellula all'altra.

Consideriamo in primo luogo la struttura della cellula nervosa. Essa riceve l'informazione sul proprio corpo cellulare e sui dendriti, prolungamenti ramificati collegati direttamente al corpo cellulare, tramite le sinapsi che realizzano i punti di continuità. La cellula nervosa integra questa informazione e quindi invia impulsi lungo il suo assone, singola fibra nervosa che si ramifica in branche terminanti in bottoni (bulbi sinaptici), il quale conduce i segnali da una sinapsi all'altra (*si veda la figura in alto a fronte*). Le cellule nervose sono generalmente riunite e raggruppate; il caso che particolarmente ci interessa è quello in cui gli assoni sono allineati fianco a fianco in un fascio, a formare una via di conduzione.

Alcune vie sono molto corte e connettono cellule nervose incluse in un raggruppamento detto nucleo; altre sono invece piú lunghe e collegano gruppi di cellule lontane parecchi millimetri. Le vie che si estendono dal cervello fino ai segmenti inferiori del midollo spinale sono lunghe circa 60 centimetri e dal midollo altre vie lunghe si dirigono verso gli organi sensoriali e motori.

L'esistenza nel cervello di vie strutturali fu scoperta parecchi anni fa; alcune di queste, grossi fasci che appaiono come fibre bianche e brillanti per la guaina mielinica che ricopre l'assone, sono visibili anche senza microscopio e la semplicità esteriore di tali vie ricopre una spettacolare complessità. Attualmente si sa che le cellule nervose e le loro interconnessioni costituiscono un insieme di strutture cosí complicate che se le si volessero raffigurare dettagliatamente bisognerebbe tracciare un tale groviglio in cui sarebbe impossibile districare i componenti.

Metodi classici di colorazione

Fu possibile individuare queste strutture quando alcuni studiosi scoprirono metodi di colorazione capaci di mettere in evidenza le strutture individuali delle cellule nervose in modo da renderne possibile l'osservazione al microscopio.

Nel 1884 Franz Nissl in Germania scoprí che il citoplasma del corpo della cellula nervosa ha una tale affinità per il blu di metilene che con questo colore, e con altri coloranti alcalini, è possibile evidenziare selettivamente il corpo cellulare. Il metodo di Nissl permette di mettere in risalto e di identificare regioni ben definite nel cervello.

Circa nello stesso periodo l'italiano Camillo Golgi sviluppò il suo metodo per la colorazione selettiva di alcune cellule nervose usando sali di argento.

Il metodo di colorazione di Golgi evidenzia molti particolari di una cellula nervosa, compresi il dendrite e l'assone. Se però il colore ha impregnato tutte le cellule in una sezione di apprezzabile spessore i dettagli possono perdersi nel fitto groviglio di fibre. Nonostante il metodo di colorazione di Golgi riesca a colorare solo una piccola percentuale di cellule, è tuttavia possibile evidenziare i dettagli di queste strutture cellulari e connessioni in una sezione di tessuto cerebrale di pochi decimi di millimetro. Per le altre tecniche di colorazione lo spessore della fettina di tessuto deve essere limitato a pochi centesimi di millimetro perché sia possibile esaminare i processi cellulari e le connessioni tra le cellule nervose. Grazie allo spessore della sezione di tessuto è possibile tracciare il processo di una cellula nervosa per un considerevole tratto, spesso per l'intera lunghezza di una corta via che connette due raggruppamenti di cellule nervose. Il vantaggio del metodo del Golgi fu subito utilizzato dallo spagnolo Santiago Ramón y Cajal la cui metodica per evidenziare le cellule nervose e le loro interconnessioni è ancora oggi fondamentalmente usata. Nel 1906 a Golgi e a Ramón y Cajal fu assegnato il premio Nobel per la fisiologia e la medicina.

Nel migliore dei casi una fotografia di una sezione colorata mostra al microscopio ottico solo frammenti e pezzetti di tessuto nervoso.

Un passo importante nel definire le strutture che costituiscono le vie del cervello è stato compiuto con l'introduzione del microscopio elettronico. Questo strumento può rivelare, con un alto grado di ingrandimento e risoluzione, i dettagli dei singoli elementi che costituiscono la complessa cellula nervosa e le interconnessioni tra le varie cellule. Il microscopio elettronico però consente un campo di osservazione molto ristretto, essendo il suo uso limitato a una sezione di tessuto sottile meno di un decimo di micron. È perciò impossibile tracciare anche la piú corta via nervosa pur riunendo insieme una serie di microfotografie. D'altra parte lo strumento è molto efficace nel delineare la struttura fine degli oggetti; esso può essere impiegato per esempio per analizzare gli elementi della giunzione sinaptica dove l'informazione passa da una cellula alla seguente.

La sinapsi è un punto focale per la ricerca della funzione del cervello. È qui che avviene l'interazione e la modificazione degli impulsi nervosi che determinano il comportamento animale. L'analisi della regione sinaptica con il microscopio elettronico ha ben presto fornito nuove informazioni sulla struttura di questo apparato di trasmissione. È stato dimostrato che, nella cellula nervosa del cervello e del midollo spinale, i bottoni da cui i segnali vengono passati attraverso la sinapsi, contengono in genere un numero di vescicole chiare (sacche piene di fluido) di diametro variabile da 300 a 600 angstrom.

Si pensa che le vescicole contengano una sostanza che serve come trasmettitore chimico del segnale. Apparentemente quando un impulso nervoso arriva a un bottone sinaptico esso libera una piccola quantità di questa sostanza la quale attraversa lo spazio sinaptico, largo da 200 a 300 angstrom, producendo un segnale che eccita o inibisce la cellula seguente.

Il tracciato delle vie nervose

Combinando il microscopio ottico con quello elettronico e utilizzando le diverse tecniche di colorazione è possibile mettere in evidenza tutti i particolari delle vie nervose. Una serie di fotografie eseguite al microscopio ottico di fibre nervose colorate con argento delinea il tracciato delle vie nervose mentre il microscopio elettronico ne delinea i particolari.

Tuttavia anche con questi mezzi è raramente possibile tracciare l'intero percorso di una via nervosa in un tessuto normale. Fortunatamente una scoperta fatta piú di un secolo fa ha dato un significato alla conduzione nel corso di indagini su preparati sperimentali.

Nel 1850 l'inglese Augustus Volney Waller osservò, nel corso di esperimenti su tessuto nervoso, che, quando una cellula nervosa è stata danneggiata, la struttura fibrosa che si trova davanti al punto lesionato, degenerava spezzan-

Una cellula nervosa con i suoi processi (*in colore*) e parti di tre altre cellule (*in nero*) sono schematizzati a sinistra. Un impulso nervoso scorre lungo un assone che si ramifica in un certo numero di branche terminali che finiscono in bottoni a contatto con i corpi e con i dendriti di altre cellule. Questi punti di contatto sono le sinapsi.

Cervello di ratto visto di lato (*a*), dall'alto (*b*) e dal basso (*c*). In ciascun disegno le linee nere marcate indicano il piano di sezione trasversale nelle successive figure.

La colorazione di Nissl evidenzia i corpi cellulari, permettendo l'identificazione dei «territori» cellulari nel cervello. Una microfotografia a basso potere di risoluzione (*a sinistra*) di una sezione trasversale di cervello di ratto eseguita da Walle J.H. Nauta e Harvey J. Karten, mostra la disposizione a strati delle cellule nella corteccia, in vicinanza della superficie del cervello. Le cellule della corteccia olfattoria (*nel rettangolo*) sono ingrandite 100 diametri nella microfotografia a destra.

La colorazione di Bielschowsky impregna soprattutto grandi fibre nervose ed è perciò efficace nel delineare grosse vie attraverso una microfotografia a basso potere di risoluzione (*a sinistra*). Tuttavia essa colora tutte le fibre in modo tale che le singole fibre non possono essere seguite con facilità. Quando una porzione della parte basale del cervello (*racchiusa nel rettangolo*) è ingrandita per esempio di 40 diametri, la complessa rete di fibre non può ormai piú essere distinta (*a destra*).

dosi in frammenti. Nel 1880 gli italiani Vittorio Marchi e Giovanni Algeri stabilirono un metodo per colorare selettivamente fibre che stavano andando in « degenerazione Walleriana ».

Secondo la loro tecnica viene colorata la guaina mielinica di tali fibre. La mielina è dapprima trattata con una soluzione di bicromato di potassio e quindi colorata con tetrossido di osmio. Ciò rende possibile tracciare una via nervosa seguendo il percorso di fibre la cui mielina stia spezzandosi. Questa tecnica tuttavia lascia ancora numerose lacune nella descrizione delle vie nervose in quanto essa non mostra quelle parti di fibre che non sono mielinizzate e che comprendono non solo le branche assoniche terminali con i loro bottoni ma anche molti dei sottilissimi rami assonici privi di guaina mielinica.

Importanti progressi furono compiuti negli anni '40 e '50, quando il perfezionamento nella colorazione di fibre nervose con l'argento consentí l'identificazione di materiale in degenerazione all'interno delle fibre. Ramón y Cajal e il tedesco Max Bielschowsky, verso la metà del secolo hanno sviluppato tecniche di impregnazione con argento metallico per colorare le fibre nervose, ma per parecchi anni questi metodi sono stati applicati solo allo studio di tessuti normali. L'applicazione di questo metodo di impregnazione a fibre degenerate ha permesso di tracciare mappe di assoni scarsamente o non mielinizzati e di bottoni sinaptici. La tecnica piú in uso è stata sviluppata da Walle J.H. Nauta in collaborazione con Paul A. Gygax dapprima in Svizzera e poi al Walter Reed Army Institute for Research a Washington. La sezione di tessuto viene pretrattata con permanganato di potassio e quindi con nitrato di uranile prima di essere impregnata con una soluzione di nitrato di argento. Quindi essa è trasferita in una soluzione argento-ammoniacale e infine ridotta in una soluzione diluita di formalina, acido citrico e alcol. La tecnica non solo è efficace per le fibre degenerate, ma sopprime anche l'affinità delle fibre normali per l'argento; per tale motivo risulta facile identificare e quindi tracciare il percorso di fibre che si stanno spezzettando. Lo sviluppo di questo conveniente metodo per tracciare il percorso di vie nervose mediante esperimenti su animali, ha permesso un notevole sviluppo nella esplorazione della struttura anatomica del sistema nervoso in questi ultimi dieci anni. La tecnica è stata in seguito perfezionata da molti ricercatori, me stesso compreso in collaborazione con Robert Fink.

Questa microfotografia eseguita al microscopio elettronico ingrandisce una parte della corteccia olfattoria di 24 000 diametri. Nello schema sono raffigurati alcuni componenti presenti nella parte della microfotografia racchiusa in bianco. Le frecce indicano la direzione di trasmissione. Le cellule gliali sono un supporto per le cellule nervose.

Le vie olfattorie

Per illustrare il procedimento usato e la natura della nostra scoperta descriverò il mio lavoro sul sistema olfattorio del cervello. In tutti i vertebrati, uomo compreso, il sistema olfattorio sembra avere lo stesso tipo di organizzazione. Le sostanze odorose che si trovano nell'aria sono percepite da cellule recettrici situate nella parte superiore della cavità nasale. Queste cellule, che sono parecchi milioni, pare siano gli unici recettori sensitivi periferici costituiti da vere e proprie cellule nervose. Ricevuto lo stimolo, gli assoni cellulari convogliano i risultanti impulsi nervosi alla prima stazione, il bulbo olfattorio del cervello. Questo è una complessa struttura da cui nella zona di uscita, parecchie migliaia di cellule mitrali (fatte come la mitra di un vescovo), inviano messaggi ai centri olfattori situati in altre zone del cervello. Altri tipi di cellule sono presenti a formare intricati circuiti nel bulbo e si pensa che il messaggio ricevuto dai recettori olfattori sia notevolmente modificato nel bulbo prima di essere trasmesso alle altre regioni del cervello.

Il sistema olfattorio del ratto ha recettori cellulari nella cavità nasale che, stimolati da sostanze odorose, trasmettono messaggi alle cellule mitrali nel bulbo olfattorio. Gli assoni delle cellule mitrali si dirigono verso la corteccia olfattoria comprendente il tubercolo olfattorio. L'autore ha trovato che le cellule del tubercolo olfattorio comunicano con le cellule della sostanza innominata. Queste potrebbero essere le stesse cellule i cui assoni sembrano estendersi dalla sostanza innominata alle cellule granulari del bulbo olfattorio, il circolo tratteggiato indica questa probabilità. La possibilità di tale connessione suggerisce la esistenza di un anello di retroazione che modifica gli impulsi per il sistema olfattorio. Altre vie possono modificare le risposte.

La via olfattoria è tracciata usando la tecnica dell'argento metallico che colora preferenzialmente assoni degenerati di cellule mitrali distrutte da una lesione nel bulbo olfattorio (*a sinistra*). Gli assoni diffondono sulla superficie inferiore dell'emisfero cerebrale (*linea tratteggiata*) e sembra che terminino proprio sotto la superficie (*a destra*). Il rettangolo grande indica la posizione della microfotografia a destra a pagina 101; quelli piccoli indicano le posizioni delle microfotografie a fronte.

Tralasciamo per il momento le strutture anatomiche del bulbo e soffermiamoci sulle fibre nervose che si dipartono dalle cellule mitrali. Il nostro principale obbiettivo nello studio del sistema olfattorio è scoprire dove terminino le vie nervose. In primo luogo abbiamo studiato dettagliatamente le terminazioni della via mitrale, oggetto di numerose controversie. Abbiamo distrutto in un ratto le cellule mitrali del bulbo olfattorio mediante un intervento chirurgico, e dopo un tempo di sopravvivenza di tre o quattro giorni, abbiamo sacrificato l'animale per condurre un esame microscopico delle fibre che si dipartono da queste cellule. Le fibre, impregnate con argento secondo la tecnica Nauta-Gygax, mostrano una traccia di degenerazione individuata da una linea di frammenti simili a gocce e collegati da ponti sottili, come potrebbe essere l'aspetto di una fibra disintegrata. Osservando una serie di sezioni fatte consecutivamente nel cervello, si possono seguire gli assoni delle cellule mitrali degenerati che formano una via nervosa, il « tratto olfattorio », visibile a occhio nudo come una curva attorno al tubercolo olfattorio. La via si estende, e diffonde a ventaglio le fibre disintegrate molto al di sopra del mantello corticale della parte basale del cervello.

In questa regione, dove termina la via nervosa, si trovano particelle nere d'argento sparpagliate apparentemente a caso (*si veda la microfotografia in alto a destra*) e ci si domanda se esse rappresentino bottoni sinaptici degenerati. Le particelle osservate al microscopio elettronico non mostrano caratteristiche tali da poter essere identificate. In collaborazione con Alan Peters dell'Università di Boston abbiamo osservato al microscopio elettronico delle sezioni di questo tessuto impregnate con argento. Le microfotografie svelarono che molte delle grandi e più o meno sferiche particelle erano senza dubbio bottoni sinaptici degenerati. Essi possono essere identificati come tali per la densità del materiale di cui sono costituiti, per il loro aspetto raggrinzito e deformato e il loro contenuto di vescicole, o per la loro associazione con una membrana specializzata intorno alla sinapsi. Al microscopio ottico il bottone degenerato appare solo come una particella nera a causa della bassa risoluzione, ma al microscopio elettronico, quando lo spessore della sezione è solo un quindicesimo di micron, possiamo osservare una piccola porzione del bottone, da circa un quindicesimo a un trentesimo del suo diametro totale. Si osserva che il bottone degenerato è solo parzialmente riempito dall'argento e possono essere discriminati dettagli della sua struttura (*si veda la figura in alto nella pagina seguente*). Le microfotografie fatte al microscopio elettronico hanno permesso di scoprire che le particelle nere più piccole trovate nella regione dove terminano le vie delle cellule mitrali, sono frammenti di branche terminali di assoni degenerati.

Seguiamo ora il percorso di una via nervosa che si diparte dal tubercolo olfattorio. Creando sperimentalmente lesioni superficiali nel tubercolo con una piastra d'argento scaldata a 70° C, abbiamo potuto seguire il percorso di fibre degenerate fino a limitate regioni del talamo e dell'ipotalamo. Abbiamo inoltre rilevato che questa via termina nella regione detta sostanza innominata, vicina e apparentemente in comunicazione con l'ipotalamo. Recentemente Joseph L. Price e Thomas P.S. Powell dell'Università di Oxford si sono particolarmente interessati a questa scoperta. Nel corso di un loro lavoro hanno osservato che gli assoni di alcune grandi cellule nervose della sostanza innominata sono diretti verso il bulbo olfattorio, dove comunicano con cellule dette granulari. Da quando si è venuti a conoscenza che il tubercolo olfattorio è collegato direttamente al bulbo olfattorio da una via che si diparte dal bulbo, sembra possibile che messaggi di ritorno possano essere inviati dal terminale del tubercolo al bulbo stesso tramite un collegamento nella sostanza innominata. Se così fosse, potremmo supporre l'esistenza di un anello di retroazione in cui l'energia che si origina dal bulbo stesso e si propaga intorno attraverso il tubercolo e la

Una prova della via tracciata nell'illustrazione in basso nella pagina a fronte è in questa microfotografia di una sezione di cervello di un animale sacrificato cinque giorni dopo una lesione al bulbo olfattorio. Gli assoni degenerati colorati all'argento sono messi in evidenza come frammenti simili a gocce raggruppate linearmente (*A, in basso*). Le particelle sferiche, colorate all'argento in uno strato più profondo (*B, in alto*), sono invece state identificate come branche assoniche degenerate in bottoni sinaptici.

Un bottone degenerato della corteccia ingrandito 25 000 diametri in una microfotografia eseguita al microscopio elettronico (*a sinistra*). L'aumento di densità e deformazione sono segni di degenerazione. Analoghi bottoni in una sezione colorata all'argento (con un tempo di sopravvivenza di 3 giorni) contengono notevoli accumuli di argento (*a destra*), la qual cosa costituisce una prova del fatto che molte particelle sferiche nelle microfotografie sono bottoni degenerati di assoni di cellule mitrali.

Sezione dopo una lesione uguale a quella rappresentata nella microfotografia in basso a pagina 102, fatta 36 ore dopo una lesione nel tubercolo. La microfotografia eseguita al microscopio elettronico mostra un dendrite circondato da un bottone degenerato scuro, e da numerosi bottoni normali. Il bottone degenerato deve appartenere a un assone proveniente dal tubercolo olfattorio danneggiato. Gli studi indicano che la via del tubercolo olfattorio è rappresentata da bottoni con grandi vescicole che rendono «simmetrico» il contatto con i dendriti delle cellule della sostanza innominata.

sostanza innominata, ritorni a modificare l'attività del bulbo. Per rispondere a questo interrogativo bisogna scoprire se le fibre del tubercolo sono in contatto con i dendriti delle stesse cellule della sostanza innominata che inviano messaggi al bulbo. Qualunque possa essere l'esatta organizzazione di queste connessioni sembra certo che il cervello moduli la trasmissione dell'informazione olfattoria rapidamente come a livello della prima sinapsi della via olfattoria.

La sostanza innominata è ricca di grossi dendriti che l'attraversano in varie direzioni. L'analisi al microscopio elettronico mostra che questi dendriti sono coperti da bottoni di molti tipi diversi, variando nella forma delle loro vescicole, nel tipo delle loro membrane sinaptiche e in altri particolari. Studi eseguiti su cellule di altre regioni del cervello, hanno dimostrato che un dato tipo di bottone sinaptico rappresenta una determinata via nervosa. Ciò suggerisce che la presenza di piú di un tipo di bottone su un dendrite stia a significare la ricezione del segnale da piú di una sorgente o via nervosa. In breve, i dendriti della sostanza innominata possono ricevere informazioni non solo dal tubercolo olfattorio ma anche da parecchie altre regioni del cervello.

Il nostro primo passo nella ricerca dell'esistenza di un anello di retroazione fu fatto per tentare di identificare il tipo di bottone che rappresentasse la via proveniente dal tubercolo olfattorio. Per fare ciò avevamo bisogno di combinare le informazioni forniteci dal

Corteccia olfattoria del cervello vista in microfotografie di sezioni colorate con il metodo di Golgi (*a sinistra*) e con il metodo all'argento metallico (*a destra*) e colorate per contrasto con cresilvioletto. La colorazione di Golgi (*nero*) permette di individuare poche cellule con i loro processi in una sezione di tessuto. La sezione di Golgi, preparata da Enrique Ramón-Moliner, mostra che alcuni dendriti (processi recettori di impulsi) di grandi cellule corticali (*al centro*) si allungano fino alla superficie inferiore del cervello (*in basso*). Il metodo di colorazione dell'argento riducente mette in evidenza gli assoni, o fibre nervose (*marrone scuro*), che sono degenerati poiché i loro corpi cellulari sono stati distrutti da una lesione sperimentale nel bulbo olfattorio. Cinque giorni dopo la lesione è stata preparata la sezione di cervello. Gli assoni degenerati si incrociano vicino alla superficie (*in basso*). Sopra vi è una zona di degenerazione estrema. In alto vi sono grandi cellule corticali.

Cellule e processi nervosi colorati in giallo in una microfotografia di una sezione colorata all'argento della sostanza innominata, area in cui terminano gli assoni che provengono dal tubercolo olfattorio. Il quadrato mette in evidenza un dendrite.

Un'analoga sezione della sostanza innominata preparata (dopo un tempo di sopravvivenza di due giorni) seguendo una lesione nel tubercolo olfattorio, che è parte della corteccia olfattoria. Le particelle argentate (*nero*) indicano i bottoni degenerati i cui corpi sono stati distrutti dalla lesione. La microfotografia a pagina 100 riproduce una regione simile a quella nel quadrato.

microscopio ottico e da quello elettronico; per eliminare le ambiguità l'esame fu effettuato su un unico campione di tessuto. Per prima cosa abbiamo tagliato una sezione semisottile e l'abbiamo colorata per l'osservazione al microscopio ottico. Quindi abbiamo usato parte dello stesso pezzo di tessuto per tagliare una sezione sufficientemente sottile per il microscopio elettronico. Il metodo usuale di preparare sezioni sottili di tessuto cerebrale, che è estremamente molle, è di indurire il tessuto con un fissativo come la formalina e l'alcol, che preserva la naturale struttura dei tessuti, e quindi congelarlo o includerlo in un materiale solido in modo da poter tagliare delle fettine sottili con il microtomo. Ho scoperto che la tecnica della impregnazione con argento per identificare fibre degenerate e bottoni può essere applicata a un tessuto incluso in plastica, il che è essenziale per tagliare fettine ultra sottili.

Dopo aver esaminato una sezione semisottile, spessa circa 5 micron, al microscopio ottico per ottenere una fotografia ben definita della struttura della fibra in degenerazione, abbiamo ridotto il blocco da cui la sezione era stata tagliata in modo tale che questo fosse limitato all'area di interesse. Da questo blocco quindi abbiamo tagliato sezioni spesse circa un quindicesimo di micron per l'esame al microscopio elettronico.

Tale analisi ha identificato il tipo di bottone coinvolto nella via del tubercolo olfattorio. Probabilmente il bottone è di un tipo che compie un contatto sinaptico « simmetrico » dato che la sua membrana apparentemente si unisce alla membrana della cellula recettrice che le si affaccia davanti attraverso lo spazio sinaptico.

Identificazione delle cellule recettrici

Dobbiamo ancora risolvere il problema dell'identificazione delle cellule della sostanza innominata che ricevono i messaggi provenienti dal tubercolo olfattorio. Le cellule recettrici sono le uniche che inviano i loro assoni al bulbo olfattorio, completando cosí l'anello di retroazione. È molto difficile identificare le cellule stesse, a causa del ristretto campo di osservazione del microscopio elettronico. Nella microfotografia elettronica possiamo osservare solo una parte di un dendrite; la fotografia non è sufficientemente ampia per mostrare il corpo cellulare a cui il dendrite è attaccato.

Fortunatamente è stata recentemente sviluppata una tecnica che permette di risolvere tale problema. Essa è basata sull'uso del metodo di colorazione di Golgi, che mostra i vari processi di una cellula nervosa, sia i dendriti che il corpo cellulare e gli assoni. William Steel dei National Institutes of Health e Theodor Blackstad dell'Università di Aarhus in Danimarca, hanno scoperto indipendentemente che, dopo aver impregnato una intera cellula nervosa per evidenziare i processi cellulari con argento secondo il metodo di Golgi, la massa di tessuti contenente la cellula può essere ridotta a un frammento mostrante solo una porzione della stessa che può essere sezionata per l'osservazione al microscopio elettronico. In tale maniera si può riconoscere che la porzione di dendrite osservata appartiene proprio alla cellula colorata e identificata in precedenza. Esistono alcune difficoltà nell'applicazione di questa tecnica che tuttavia può aiutare nello stabilire quali cellule della sostanza innominata siano legate alla via che si diparte dal tubercolo olfattorio.

Come Ramón y Cajal ha voluto sottolineare, « la scoperta è una funzione del metodo usato », e questa verità è ampiamente dimostrata per lo studio delle vie nervose del cervello. Oltre alla tecnica di Golgi, ai metodi di colorazione con argento perfezionati e all'uso del microscopio elettronico, che è fondamentale per questo genere di lavoro, sono validi altri metodi. Esistono tecniche che sono basate sulla fluorescenza e anche l'autoradiografia promette dei buoni risultati. Esistono ancora altre tecniche chimiche che avranno un ruolo sempre piú importante nel chiarire i sistemi neuronici.

È improbabile riuscire a tracciare mappe dettagliate dell'intero plesso nervoso del cervello con la sua innumerevole moltitudine di fibre e di contatti sinaptici. Tuttavia ci sono buone prospettive per ottenere la base di partenza di molti significativi circuiti di funzionamento e per individuare quali di queste vie contribuiscano a modificare il comportamento. Sarà per esempio interessante identificare il reticolato delle vie che originano in varie regioni del cervello e convergono a sistemi centrifughi che portano l'informazione al bulbo olfattorio. L'andamento degli ingressi da queste varie sorgenti senza alcun dubbio influenza gli impulsi in uscita dal bulbo con conseguenti effetti su quei comportamenti che sono influenzati dall'odore, cioè su moltissimi comportamenti diversi che vanno dal mangiare all'accoppiamento.

L'effetto sul comportamento dell'odore di un piatto di bocconcini piccanti, per esempio, dipende moltissimo ovviamente dal fatto di essere piú o meno affamati.

GIOCHI MATEMATICI

di *Martin Gardner*

Alcune considerazioni sulle strutture elicoidali, dal cavatappi alle molecole del DNA

Un pezzo di filo di ferro rettilineo può agevolmente introdursi in una guaina rettilinea. Lo stesso può dirsi per un filo di ferro sagomato secondo un arco di circonferenza: esso può agevolmente introdursi in una guaina avente la sua stessa curvatura. Talora, in linguaggio matematico, questa proprietà della circonferenza e della retta viene espressa dicendo che tali curve sono « autocongruenti »: un qualsiasi loro segmento può essere fatto scorrere lungo tutta la curva, da un estremo all'altro, rimanendo sempre in contatto con essa con tutti i suoi punti.

Un'elica circolare (**linea colorata**) *su una superficie cilindrica.*

È possibile sagomare un pezzo di filo di ferro in modo tale che, pur *non* assumendo la forma di un segmento di retta o di un arco di circonferenza, risulti ugualmente introducibile in una guaina che riproponga la sua stessa forma? Molti sarebbero portati a dare una risposta negativa, ma commetterebbero un grosso errore. Esiste una terza curva autocongruente: l'elica circolare. Si tratta di una curva tracciabile su una superficie cilindrica e rappresentata da una linea che si avvolge attorno alla superficie cilindrica intersecando le infinite generatrici della superficie stessa secondo un angolo costante. L'illustrazione qui accanto mostra chiaramente questa proprietà. Converrà a tale proposito ricordare che per generatrici di una superficie cilindrica si intendono tutte le rette della superficie cilindrica parallele all'asse di simmetria del cilindro; nella figura, a indica l'ampiezza costante dell'angolo formato dall'elica con una generica generatrice.

Al variare dell'angolo a, l'elica degenera in due notissime curve limite: la retta e la circonferenza. Comprimendo la curva fino ad avvicinare di molto le spire l'una all'altra, si ottiene una spirale fortemente attorcigliata; al tendere poi di a a una ampiezza di 90°, l'elica si riduce a una circonferenza. D'altro canto, al tendere di a a zero, l'elica degenera appunto in una retta.

Una qualsiasi elica, circolare o no, è una curva spaziale asimmetrica; essa differisce cioè dalla sua immagine speculare. Stabiliremo da ora in poi di indicare brevemente con elica « destra » una elica che si avvita in senso orario, cosí come avviene ne'l'avvitare una normale vite o un cavatappi. Premesso ciò, è sufficiente considerare la immagine riflessa del cavatappi in uno specchio, per constatare che la spirale che lo costituisce, esprimendoci con le parole dell'Alice di Lewis Caroll, « si avvita dall'altra parte ». L'immagine riflessa è infatti un cavatappi « sinistro ». Un cavatappi del genere si presta bene a essere considerato come un vero e proprio « oggetto misterioso » particolarmente adatto per tramare una burla a danno di qualche amico. Infatti se si invita il malcapitato a fare uso di un cavatappi « sinistro » per stappare una bottiglia, passeranno certamente vari minuti di dura lotta prima che la « vittima » si renda conto, e con stupore, che per potere usare un simile ordigno lo si deve avvitare in senso antiorario.

A parte i cavatappi, i bulloni e i dadi che sono (tranne che in alcuni casi particolari) generalmente caratterizzati da una struttura elicoidale destra, moltissime strutture elicoidali costruite artificialmente presentano un verso di avvitamento sia destro che sinistro: i bastoncini di zucchero candito, le scale a chiocciola, le funi e le gomene costituite da corde intrecciate e cosí via. La stessa varietà di « avvitamento » è rintracciabile nelle eliche coniche (curve che si attorcigliano a spirale su una superficie conica), comprese le molle delle reti dei letti e le rampe a spirale, per esempio, delle scale del Guggenheim Museum di New York.

In natura, però, non accade altrettanto! La presenza nelle forme viventi di strutture elicoidali è estremamente frequente, a partire dal piú semplice dei virus per giungere perfino a componenti del corpo umano; e quasi in ogni caso il codice genetico trasporta tutte le informazioni sufficienti per ordinare a ogni elica « il modo in cui deve avvitarsi ». Infatti, il codice genetico stesso risulta trasportato e custodito in singole macromolecole di acido nucleico caratterizzate proprio (almeno allo stadio attuale delle nostre conoscenze biochimiche) dall'essere strutturate a elica destrorsa. Inoltre, dai lavori pionieristici di Linus Pauling sulle strutture elicoidali delle molecole proteiche, si sono trovate sempre ulteriori prove a favore del fatto che qualsiasi macromolecola presente in natura presenta uno « scheletro » sagomato proprio secondo uno schema elicoidale destrorso. Sia nel caso degli acidi nucleici che delle proteine, lo scheletro delle molecole è a sua volta costituito da un groviglio di singole unità elicoidali asimmetriche e aventi lo stesso verso di avvitamento. Ognuna delle singole unità, per cosí dire, imprime un ulteriore avvitamento alla catena, nella stessa direzione, come i singoli gradini di una scala a spirale.

Negli animali a struttura corporea simmetrica, sono spesso presenti strutture elicoidali molto evidenti, che in generale sono poste simmetricamente sulle due bande del corpo e sono l'una la immagine speculare dell'altra. Le corna delle antilopi, dei montoni, delle capre e di numerosi altri mammiferi forniscono altrettanti vistosi esempi di questo fenomeno (*si veda l'illustrazione in alto*). La coclea dell'orecchio umano è una elica conica sinistrorsa nell'orecchio sinistro e destrorsa in quello destro. Una curiosa eccezione è rappresentata dal dente del Narvalo, una piccola balenottera che prospera nei mari artici. Questa eccentrica creatura è dotata di due denti entrambi localizzati sulla mascella superiore. Mentre nelle femmine tali denti restano di solito nascosti nell'osso della mascella, nel maschio ciò si verifica solo per il dente destro. Ma il dente sinistro del maschio sporge in avanti come un giavellotto e si accresce avvolto a spirale, fino a raggiungere anche un paio di metri di lunghezza: lunghezza straordinaria per un dente, dal momento che esso supera spesso la metà della lunghezza totale dell'animale! La conformazione a spirale
tesco dente procede in verso antiorario (*si veda l'illustrazione in basso*). Nei rari casi in cui entrambi i denti del cetaceo si sviluppano fino ad assumere l'aspetto di vere e proprie zanne, si sarebbe portati a supporre che il dente destro dovesse presentare un andamento spiraliforme orario. Ciò però, e in questo consiste l'eccezione, non si verifica: entrambi i denti presentano lo stesso verso di avvitamento antiorario. Gli zoologi si trovano in disaccordo per quanto riguarda l'interpretazione da dare a un tale fenomeno. Sir D'Arcy Thompson, nel suo libro *Crescita e forma*, sostiene una sua personalissima teoria: questi cetacei procedono nell'acqua tendendo a imprimere alla loro massa corporea una lieve rotazione in senso destrorso; l'inerzia della loro enorme zanna produrrebbe una coppia applicata alla base del grande dente capace di imprimere al dente stesso una orientazione antioraria durante la crescita.

Ogniqualvolta nella struttura di un qualsiasi essere, vegetale o animale, si pongono in evidenza elementi elicoidali, la specie di solito si caratterizza con un ben determinato e specifico verso di avvitamento. La cosa si verifica per le innumerevoli forme di batteri elicoidali, come pure per gli spermatozoi di tutti gli animali superiori. Il cordone ombelicale dell'uomo è un'elica triplice costituita da una vena e da due arterie, che invariabilmente si attorcigliano in senso sinistrorso. Gli esempi piú interessanti e sorprendenti sono forniti dalle eliche coniche delle conchiglie delle chiocciole e di altri molluschi. Non tutte le conchiglie a spirale hanno un ben determinato verso di avvitamento. Il guscio del Nautilo, per esempio, si avvolge a spirale mantenendosi su di un piano; cosí come una qualsiasi nebulosa a spirale, esso può percorrersi indifferentemente sia in verso orario che antiorario. Ma esistono migliaia di splendide conchiglie di molluschi che sono destrorse o sinistrorse (*si veda l'illustrazione in alto nella pagina seguente*). Alcune specie sono sempre sinistrorse, e alcune sempre destrorse. Alcune vanno in un verso in una località e nel verso opposto in un'altra. Eventuali « eccezioni » che si avvitano nel senso « sbagliato », costituiscono prede pregiate per i collezionisti.

Le corna elicoidali della pecora del Pamir hanno versi opposti.

Uno sconcertante tipo di fossile a struttura elicoidale, noto con il nome di « cavatappi del diavolo » (*Daemonelix*) fu portato alla luce nel Nebraska e nello Wyoming. Si tratta di grandi gusci spiraliformi, lunghi circa due metri, e che sono talvolta destrorsi e talvolta sinistrorsi. Per decenni i geologi restarono incerti sulle origini e sulla natura di tali fossili: erano tracce fossilizzate di vegetali ormai estinti o erano invece cunicoli elicoidali scavati da lontani antenati dei castori e da essi usati come tane? Alla fine, su tutte, prevalse la « teoria dei castori », nel senso che quelle strane strutture elicoidali si ritiene fossero la parte interna dei cunicoli scavati come tane da piccoli castori, che li sagomarono proprio a « cavatappi ».

Nel mondo vegetale le eliche sono molto comuni nella struttura dei tronchi, degli steli, dei viticci, dei semi, dei fiori, delle pigne, delle foglie e anche nella disposizione stessa nella quale le foglie e i rami si susseguono lungo il tronco. In quest'ultimo caso, è notevole il fatto che il numero delle spire da percorrersi lungo il cammino a elica che collega ogni foglia con le successive, fino a giungere a una foglia allineata, lungo una retta ideale parallela all'asse del tronco, con la foglia di partenza, risulta molto spesso essere un numero della ben nota successione di Fibonacci: 1, 2, 3, 5, 8, 13, ... (ogni termine è la somma dei due termini che lo precedono). Copiosa è la letteratura riguardante questo argomento, noto con il nome di « fillotassi » (disposizione delle foglie), che si occupa della sorprendente presenza della successione di Fibonacci in fenomeni botanici di questo genere. (Si veda per esempio *Giochi matematici* in « Le Scienze », n. 12, agosto 1969).

Le volute elicoidali degli steli delle piante rampicanti si avvitano generalmente in senso destrorso; sono però migliaia le varietà di piante rampicanti in cui compare un andamento opposto. Il caprifoglio, per esempio, si sviluppa sempre secondo volute di elica sinistrorse; i vilucchi delle siepi (famiglia di vegetali a cui appartiene anche l'*Ipomea purpurea*, i comuni convolvoli o campanellini) crescono invece seguendo uno sviluppo elicoidale destrorso. Qualora le due varietà di piante vengano fatte crescere assieme, ne risulta un appassionante e violento intreccio vegetale che ha a lungo affascinato i poeti inglesi. « L'azzurro convolvolo », scriveva Ben Jonson nel 1617, « che si avvinghia al serpentino caprifoglio ». Anche Shakespeare, nel suo *Sogno di una notte di mezza estate*, fa dire alla

La struttura elicoidale del dente del Narvalo è sempre sinistrorsa.

Tre gusci di molluschi che presentano uno schema a elica conica destrorsa.

regina Titania, innamorata di Rocchella il tessitore (che era stato trasformato in ciuco), « Dormi: ti cingerò colle mie braccia/... Al caprifoglio soave il convolvolo cosí lieve si allaccia, e cosí l'edera inanella i rugosi diti all'olmo ». Si noti che ai tempi di Shakespeare il termine « caprifoglio » indicava un generico rampicante. Poiché nel seguito il termine passò a indicare solo il caprifoglio, accadde che molti commentatori giunsero a dare una interpretazione veramente assurda al passo sopra riportato: Titania parlerebbe di un caprifoglio intrecciato con un caprifoglio! Il senso della metafora di Titania sta invece nella sua consapevolezza degli opposti versi in cui il convolvolo e il caprifoglio si sviluppano.

È evidente che la assegnazione di un verso destrorso o sinistrorso al modo in cui si « avvita » una data elica è un fatto del tutto convenzionale. Se si osserva, infatti, una vite da legno dalla parte della punta, l'osservatore vedrà la elica della vite avvolgersi in verso antiorario rispetto a se stesso; ciò evidentemente giustifica, da parte dell'osservatore, il definire quell'elica « sinistrorsa ». È però possibile adottare la convenzione opposta (cioè mettersi dal punto di vista opposto, ossia dalla parte della testa della vite).

Sull'intreccio di due eliche circolari di versi opposti era basato un ingegnoso giocattolo, fondato su una illusione ottica, in voga negli USA circa 30 anni fa. È facile costruirlo intrecciando fra loro due serpentine metalliche spiraliformi e con orientamenti opposti (*si veda l'illustrazione in basso*) che una volta intrecciate debbono essere saldate fra loro in piú do da ottenere una struttura rigida. L'illusione ottica si produce premendo l'oggetto, tenuto fra pollice e indice di entrambe le mani alle due estremità, lungo il suo asse. Spostando le mani, accade che i pollici e gli indici tendano a slittare lungo le spire, producendo la rotazione di tutta la struttura e creando da ambo i lati e con versi opposti, l'illusione ottica, estremamente comune, visibile nei classici pali a spirali bianche e rosse che costituiscono le insegne dei barbieri americani. Un tale fenomeno può ovviamente ripetersi quante volte si vuole: le spire sembra che vadano miracolosamente creando un inesauribile intreccio mobile. Poiché al livello delle conoscenze attuali il neutrino e l'antineutrino sembra che si muovano lungo una traiettoria elicoidale con versi fra loro opposti, sono portato a credere che il giocattolo di cui stiamo parlando risulti un buon modello a cui ispirarsi per visualizzare in termini macroscopici l'incessante produzione di neutrini e delle loro particelle opposte, gli antineutrini.

Il giocattolo costituito da due eliche e che suggerisce la produzione dei neutrini.

Il carattere elicoidale della traiettoria dei neutrini risulta dalla composizione dei due moti a cui sono soggetti: un moto rettilineo (alla velocità della luce) e un moto rotatorio o « spin ». Traiettorie elicoidali di tale natura possono rintracciarsi nel moto di numerosi oggetti inanimati e di vari esseri viventi: un qualsiasi punto di una elica di un natante in movimento su un piano, uno scoiattolo che sale o scende attorno al tronco di un albero, le code di una singolare specie di pipistrelli messicani che si avvolgono a spirale in verso antiorario quando gli animali vanno emergendo dalle loro tane; i caratteristici solchi scavati dai vortici, il cadere dell'acqua nelle condutture forzate, i cicloni e migliaia di altri fenomeni naturali, forniscono altrettanti esempi di percorsi ad andamento elicoidale ottenuti con la composizione dei due movimenti di cui sopra.

Gli scrittori hanno trovato i moti elicoidali estremamente utili per esporre in un linguaggio metaforico vari concetti. Il progresso scientifico è spesso rappresentato metaforicamente da un tracciato a spirale conica percorso a ritroso: le spire vanno sempre piú allargandosi proprio a indicare il continuo progredire delle scienze nel campo del del sapere, conoscenze destinate ad ampliarsi sempre piú, proprio come avviene a ogni spira dell'elica conica rispetto alla precedente. La stessa spirale, triste simbolo in questo caso, vortice senza fondo in cui ogni individuo e la stessa umanità va scivolando, è anche stata usata come simbolo del pessimismo e della disperazione piú tetra. A chiusura del libro *Advertisements for Myself*, l'autore Norman Mailer porta proprio la metafora appena accennata. « Sono ormai alla fine della strada? », si chiede. Il tempo per Mailer è rappresentato da un'elica conica di acqua che si abbatte su di lui come un fiume cosmico, che lo fa turbinare « dentro una spirale piena di acqua illuminata dalle stelle ».

E ora un semplice rompicapo riguardante le spirali, la cui soluzione troverete il mese prossimo: si consideri una insegna di barbiere, costituita da un cilindro rotante su cui sono dipinte tre eliche, rispettivamente blu, bianca e rossa. Il cilindro sia alto 2 metri. La spirale rossa interseca le direttrici del cilindro (le rette verticali appartenenti alla superficie del cilindro) secondo un angolo costante di 60°. Quanto è lungo il tratto di spirale rossa disegnato sull'insegna?

Il problema proposto il mese scorso ha la seguente soluzione. La probabilità che due carte scelte a caso da un insieme di due carte rosse e di due carte nere siano dello stesso colore è 1/3. Elencando infatti le 24 permutazioni ugualmente probabili delle quattro carte e prendendo in esame due posizioni qualsiasi (per esempio la seconda e la quarta carta), si trovano infatti otto casi in cui esse hanno lo stesso colore. Un modo per vedere che questa probabilità è 8/24, cioè 1/3, è quello di considerare una delle due carte scelte. Supponiamo che si tratti di una carta rossa; delle tre carte restanti una
probabilità che la seconda carta scelta sia rossa è 1/3. Naturalmente lo stesso argomento si applica al caso in cui la prima carta sia nera. Molti pensano che le possibilità siano pari, ma in realtà esse sono due a uno in favore delle carte aventi colore diverso.

Il gioco di Conway pubblicato nel numero di maggio (e la successiva analisi degli automi cellulari pubblicata in luglio) continua a interessare i nostri lettori che ci comunicano continuamente nuove scoperte e osservazioni.

L'Istituto tecnico femminile statale « Principessa di Piemonte » di Roma ha addirittura utilizzato il gioco di Conway nei suoi corsi per programmatori elettronici; la preside dell'istituto ci ha infatti inviato i tabulati ottenuti con un calcolatore IBM 1130, contenenti la soluzione di alcuni dei problemi proposti e anche una breve relazione della tecnica di programmazione adottata. I tabulati contengono anche l'evoluzione di due nuove configurazioni che differiscono per un solo individuo e che sono state rispettivamente battezzate collana grande e collana piccola (*si veda la figura in questa pagina*). La prima si stabilizza dopo sette mosse in due alveari, mentre la seconda dà luogo, dopo cinque mosse, a una barca lunga.

La relazione lamenta l'impossibilità di seguire l'intera evoluzione del pentomino *R* data l'esiguità della scacchiera disponibile (una matrice 20 × 20). Siamo però in grado di risolvere finalmente questo problema e anche quello del destino della riga 5-5-5-5-5-5-5 (sette insiemi di cinque organismi disposti sulla stessa verticale e separati ciascuno da una casella vuota).

Il pentomino *R* raggiunge lo stato stabile dopo 1103 mosse; nel frattempo ha prodotto sei alianti che si sono allontanati verso l'esterno. I detriti rimasti sono quattro lampeggiatori, una nave, una barca, una pagnotta, quattro alveari e otto blocchi. Questo risultato è stato ottenuto da ricercatori di diversi centri di calcolo americani, parecchi dei quali hanno anche svelato il destino della riga 5-5-5-5-5-5-5. Questa riga si

La collana piccola (sopra) e la collana grande (sotto).

stabilizza dopo 323 mosse e produce quattro semafori, otto lampeggiatori, otto pagnotte, otto alveari e quattro blocchi, per un totale di 192 bit. È interessante notare che la configurazione finale ha una duplice simmetria orizzontale e verticale giacché conserva la simmetria della configurazione iniziale: la simmetria iniziale non può infatti essere modificata durante la « vita » di una configurazione.

Alcuni lettori ci hanno segnalato che le configurazioni a riga del tipo *n-n-n-*... diventano interessanti solo quando *n* è almeno 5. Le righe 1-1-1-... e 2-2-2-... svaniscono infatti immediatamente. La riga 3-3-3-... è semplicemente una riga di lampeggiatori, mentre la riga 4-4-4-... diventa una riga di alveari stabili alla seconda mossa. In particolare, una riga di otto organismi adiacenti si stabilizza in quattro blocchi e quattro alveari, mentre la riga 5-5-5 si stabilizza in quattro blocchi e due lampeggiatori.

Infine, nella figura pubblicata nel numero di giugno e che mostrava una astronave « fuori peso » con due scorte è stato commesso un errore: la scorta superiore avrebbe dovuto essere separata di tre caselle dalla astronave.

sola è rossa e qu